의학박사 학위논문

Molecular profiling of adenocarcinoma of gastroesophageal junction compared to esophageal and gastric adenocarcinoma

위식도경계부 선암의 위선암 및 식도선암과의 분자생물학적 비교 분석 연구

2017 년 08 월

서울대학교 대학원

의학과 외과학

서 윤 석

A thesis of the Degree of Doctor of Philosophy

위식도경계부 선암의 위선암 및 식도선암과의 분자생물학적 비교 분석 연구

Molecular profiling of adenocarcinoma of gastroesophageal junction compared to esophageal and gastric adenocarcinoma

August 2017

The Department of Surgery,

Seoul National University

College of Medicine

Yun-Suhk Suh

ABSTRACT

Molecular profiling of adenocarcinoma of gastroesophageal junction compared to esophageal and gastric adenocarcinoma

Yun-Suhk Suh
Department of Surgery
The Graduate School
Seoul National University

Introduction: Biologic understanding of adenocarcinoma of gastroesophageal junction (AGEJ) and similarity to gastric or esophageal adenocarcinoma has been long standing controversial issue. The purpose of our study is to evaluate molecular characteristics of AGEJ compared to esophageal (EAC) or gastric adenocarcinoma using next generation sequencing NGS data of the Cancer Genome Atlas (TCGA) and Seoul National University (SNU) cohorts.

Methods: We retrieved NGS data of esophageal adenocarcinoma (EAC, n=78), adenocarcinoma of gastroesophageal junction or cardia (GEJ/cardia, n=48) and gastric adenocarcinoma located at fundus or body of the stomach (GCFB, n=102) from TCGA cohort. For SNU cohort, whole exome and transcriptome sequencing were carried out for each pair of tumor and corresponding normal gastric mucosae of AGEJ II (n=16 pairs), AGEJ III (n=16 pairs) and upper third gastric adenocarcinoma (UT, n=14 pairs). Class prediction model was developed using Bayesian compound covariate predictor (BCCP) with Leave-one-out cross validation between EAC and GCFB of TCGA cohort, and tested for GEJ/cardia tumors from TCGA and all tumors from

SNU cohort.

Results: The class prediction model using 400 differentially expressed classifier genes (90.2% of sensitivity and 89.7% of specificity) showed a spectral transition of clusters between EAC-like and GCFB-like group without any entirely distinguishable cluster. Using 0.4535 of BCCP score as a cut-off value, 68.8% of GEJ/Cardia of TCGA cohort and AGEJ II of SNU cohort were identified as GCFB-like group. AGEJ III of SNU cohort consisted of 93.7% of GCFB-like adenocarcinoma, and there was no significant relationship between involvement of GEJ and molecular classification of AGEJ III. EAC-like group was significantly related to differentiated and intestinal type, and showed significantly amplified copy number of ERBB2 compared to GCFB group. Reverse phase protein array and tissue microarray revealed significant overexpression of EGFR and ERBB2 in EAC-like than GCFB-like group. Drug response analysis of lapatinib from Cancer Cell Line Encyclopedia database demonstrated significantly lower half maximal inhibitory concentration for EAC-like than GCFB-like.

Conclusions: Molecular classification of AGEJ using BCCP with 400 classifier genes demonstrated that GEJ/cardia in TCGA cohort and AGEJ II in SNU cohort were a combination of 31.2% of EAC-like group and 68.8% of GCFB-like group. EAC-like group was significantly related to differentiated, intestinal type and shows significant copy number amplification of ERBB2 and overexpression of ERBB2 and EGFR. EAC-like group can be a promising target for EGFR and ERBB2 tyrosine

kinase inhibitor.

Keywords: esophagogastric junction, stomach neoplasm, esophageal neoplasm, genomics, sequencing

Student number: 2012-31122

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Adenocarcinoma of gastroesophageal junction (AGEJ) has long-lasting controversial issues for its classification or treatment strategy compared to esophageal or gastric adenocarcinoma(1-5). For classification, the Siewert classification, one of the most common clinical classification, has classified AGEJ as distal esophageal, true cardia, and subcardia cancers, but the other famous classifications, latest AJCC TNM classification or Japanese classification of gastric carcinoma, classified AGEJ with different criteria(1, 6, 7). The 8th edition of AJCC TNM classification regarded AGEJ as esophageal adenocarcinoma or gastric adenocarcinoma based on only distance between tumor epicenter and gastroesophageal junction (GEJ)(7). However, latest Japanese classification of gastric carcinoma used both distance criteria between tumor epicenter and GEJ and how much portion of tumor involved esophagus or stomach, which have great influence on treatment strategy(6). Our previous study proposed that, in terms of postoperative prognosis, AEJ arisen with the stomach should be considered as a part of gastric cancer irrespective of GEJ involvement(5). There have also a series of controversial issues regarding appropriate treatment for AGEJ. Because of the location of AGEJ between chest and abdomen, AGEJ has been in the middle of discussion about which approach between transthoracic approach or transhiatal approach would be more appropriate. Previous well-designed phase III clinical trials reported that , for Siewert type I, extended transthoracic approach which was usually

considered for esophageal cancer showed an ongoing trend towards better 5-year survival, but, for patients with Siewert II or III, transthoracic approach did not improve survival and led to increased morbidity compared with transhiatal approach which was usually considered for gastric cancer (4, 8-10) However, for Siewert II, still there have been endless debates about the extent of mediastinal/supradiaphragmatic or other extended lymphadenectomy(11-14). Considering complete mediastinal lymphadenectomy requires transthoracic approach like esophageal cancer, it is also difficult to answer for debate whether AGEJ should be managed as a part of esophageal or gastric cancer in the field of surgical treatment for AGEJ, even after several clinical trials. In terms of adjuvant chemotherapy, well-designed clinical trials have reported survival benefit of surgery plus adjuvant chemotherapy (S-1 only or Capecitabine plus Oxaliplatin) compared to surgery alone(15, 16). Considering total gastrectomy has been usually performed for advanced AGEJ, deterioration in nutritional status and functional deficit after surgery may lead to inadequate dose or cycles of postoperative adjuvant chemotherapy. However, in previous famous clinical trials including those two pivotal trials, subgroup analysis for AGEJ was not reported, and it is also difficult to predict drug response of AGEJ because tumor biology has not been comprehensively explained compared to esophageal or gastric adenocarcinoma yet(17, 18).

Consequently, more fundamental questions of biologic entity have been continuously raised, especially about whether AGEJ should be

understood as a part of esophageal adenocarcinoma or gastric adenocarcinoma. However, a few previous studies for biologic entity of AGEJ used to describe ambiguous location information of cardia cancer or be evaluated without appropriate comparative analysis, which still led to inconclusive debate of AGEJ (19-21). In the past, the incidence of esophageal or gastric adenocarcinoma as a control group showed large epidemiologic difference between the West and the East (high incidence of esophageal with low incidence of gastric adenocarcinoma in the West, and low incidence of esophageal with high incidence of gastric adenocarcinoma in the East), even though that of AGEJ now shows worldwide rapid increasing incidence pattern also in eastern countries (22-25). This epidemiologic difference makes comparative analysis among AGEJ, esophageal and gastric adenocarcinoma more difficult as we reported previously(18). There is also another conflicting issue about different characteristics for AGEJ itself between the East and the West. According to the traditional Siewert classification, AGEJ in the East has been known to have extremely low prevalence of Siewert type I and much more common type III than that in the West, which means that tumor involvement of distal esophagus by AGEJ was expected to be much less in the East(26-28). Therefore, it becomes much more difficult to perform detailed clinicopathologic analysis of each subtype of AGEJ compared to esophageal and gastric adenocarcinoma(29, 30).

In the era of molecular biology, molecular characteristics by gene expression pattern was successfully introduced for not only understating

disease entity but also new molecular classification and related treatment strategy (31-34). Regarding AGEJ, several comparative biologic investigations using conventional laboratory experiments including mutation analysis, amplification, or immunohistochemistry also have reported that AGEJ might have distinct pathological entities from gastric/esophageal adenocarcinoma and be linked to multiple genetic alterations (35-37). However, those results are still inconsistent to understand biologic similarities or differences of AGEJ compared gastric or esophageal adenocarcinoma using only one or a few molecular factors. Since 2011, molecular classification using genomic technology has been introduced in gastric cancer to distinguish epidemiologic or histologic distinction by gene expression data(32). For AGEJ, limited studies reported several differentially expressed gene expression between cardia and noncardia cancer, but not enough to understand biologic characteristics compared to esophageal or gastric adenocarcinoma (38, 39). Even in a study using targeted deep sequencing, there was a limitation not to compare AGEJ to both gastric and esophageal adenocarcinoma simultaneously, and any clinical significance was not introduced after comparison(40). Recently, the Cancer Genome Atlas (TCGA) reported comprehensive molecular classification for gastric cancer and esophageal adenocarcinoma (41, 42). Unfortunately, these world-wide large molecular analysis also do not have detailed location information or traditional clinical classification of AGEJ or cardia cancer, and study population is largely deviated to

Western society (about 25.7% of East Asian samples) even though there was significant epidemiologic difference between the East and West. Therefore, it is still unclear to investigate similarity or difference of AGEJ compared to gastric or esophageal adenocarcinoma with significant clinical relevance. However, if Eastern data including detailed location information of AGEJ will be integrated, we may expect that this large comprehensive next-generation sequencing database could be more useful supportive source to overcome several long-standing hurdles for analysis among AGEJ, esophageal and gastric adenocarcinoma.

In this study, we hypothesized that AGEJ may 1) have entirely similar characteristics to esophageal or gastric adenocarcinoma, 2) be a certain combination of esophageal or gastric adenocarcinoma, or 3) have entirely unique molecular biologic characteristics distinct from esophageal or gastric adenocarcinoma. The purpose of our study is to reveal molecular characteristics of AGEJ compared to esophageal and gastric adenocarcinoma using next-generation sequencing data of TCGA and Seoul National University (SNU) cohort.

MATERIALS AND METHODS

1. Study population of TCGA cohort

For TCGA cohort, we reviewed database of the Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/tcga/), and retrieved data of mRNA expression, somatic mutation, insertion/deletions, copy number alteration, and reverse phase protein array (RPPA) of pure esophageal adenocarcinoma (EAC), adenocarcinoma of gastroesophageal junction or cardia (GEJ/cardia) and pure gastric adenocarcinoma located at fundus or body of the stomach (GCFB) (Figure 1).

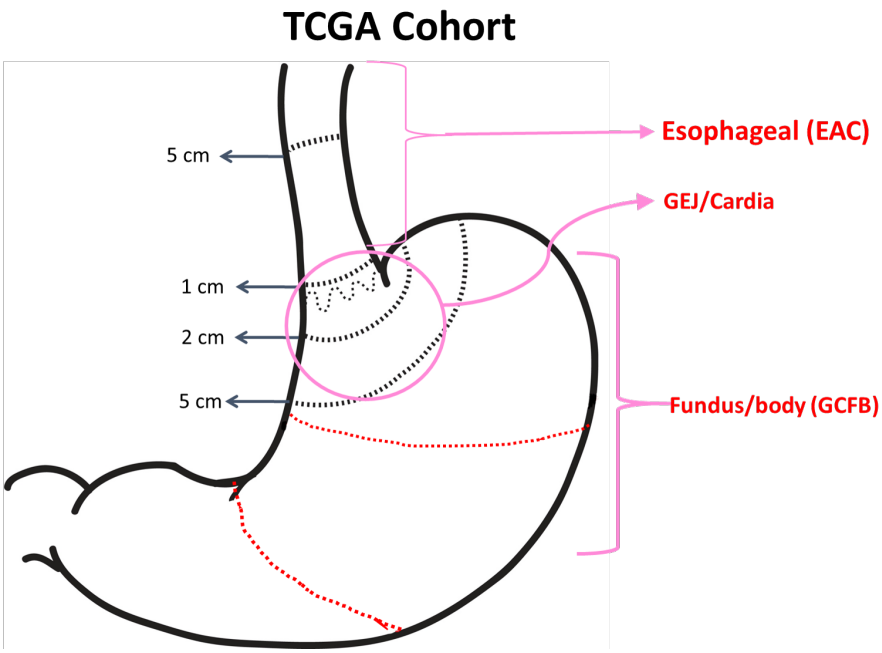


Figure 1. Anatomical distribution of study population from TCGA cohort

2. Study population of SNU cohort

For SNU cohort, we reviewed fresh frozen tissue repository database including clinicopathologic information for AGEJ and adenocarcinoma of upper third of the stomach between 1999 and 2015 at lab of gastric cancer biology, cancer research institute, SNU (Figure 2).

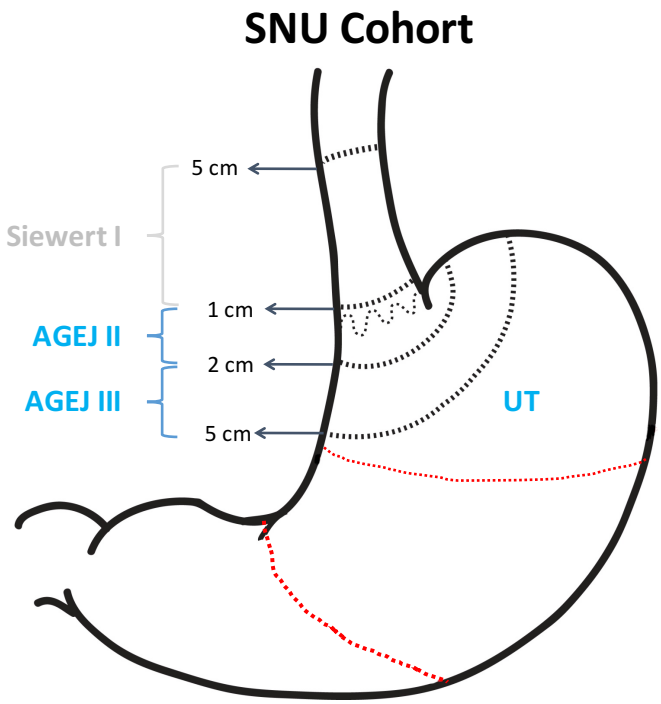


Figure 2. Anatomical distribution of study population from SNU cohort.

This fresh tissue repository was approved by the Institutional Review Board of SNU Hospital (IRB No: H-0806-072-248). Patients who had other primary malignancy, recurrent adenocarcinoma or remnant gastric cancer at the time of initial diagnosis were excluded. AGEJ and adenocarcinoma of upper third of the stomach in SNU cohort were classified using a distance criteria from the gastroesophageal junction;

AGEJ II was defined as tumor with an epicenter located within 1 cm oral and 2 cm aboral from the gastroesophageal junction, which is the same as traditional definition of Siewert type II cancer (1). AGEJ III was defined as tumors with an epicenter located within 2-5 cm aboral from the gastroesophageal junction irrespective of the involvement of gastroesophageal junction. The remaining upper one-third gastric adenocarcinoma except for AGEJ II or AGEJ III was defined as UT. All available tumors classified as AGEJ II were reviewed and prepared for next-generation sequencing. In terms of AGEJ III and UT, the same number of samples were reviewed out of the latest samples. Pathologic stage was diagnosed by the 7^{th} AJCC TNM classification (43). For pathologic analysis, papillary, well-differentiated and moderately-differentiated types were classified as a differentiated group, and poorly-differentiated, mucinous, poorly cohesive cell types were classified as an undifferentiated group (44). Regarding microsatellite instability in SNU cohort, fragment analysis was used for which tumor and normal tissue were compared at 5 point of basepair after polymerase chain reaction using following 2 primers. Primer No.1 consisted of BAT26 (116bp) and BAT25 (148bp), and primer No.2 consisted of D5S346 (96-122bp), D17S250 (146-165bp) and D2S123 (144-174bp).

This study protocol was approved by the Institutional Review Board of Seoul National University Hospital (IRB No: H-1501-027-639).

3. Nucleic acid processing, qualification of SNU cohort

Each fresh frozen tumor and corresponding normal gastric mucosa was retrieved by about 2 x 2 x 1 mm^3 from fresh tissue repository of SNU cohort. DNA was extracted using Qiagen DNA extraction kit with Spin-Column protocol (Qiagen, Venlo, Netherlands). Extracted DNA was quantified of a minimum A260/280≧1.7 and amount of dsDNA ≧3.0ug using the QUBIT HS dsDNA assay (Life Technologies Gaithersburg, MD, USA). The isolation of RNA was performed in Eppendorf Tubes 5.0 mL in accordance with the protocol provided by the manufacturer of TRIzol [User manual TRIzol® Reagent (www.invitrogen.com)]. For lysis, 1 mL of TRIzol was added for every 4 mm3 of fresh tissue. The transfer of 1 mL of the starting material into each tube was followed by the addition into each tube of 200 µL of chloroform, according to the TRIzol protocol. For precipitation of RNA from the aqueous phase 0.5 mL of isopropanol and for the following wash step 1 mL of ethanol (75 %) were used. The RNA precipitation and wash steps were carried out at 12,000 x g in the 5.0 mL tubes. The resulting RNA pellet was then resuspended in 50 µL of DEPC-treated water. Using NanoDrop™ 1000 (Thermo Scientific), OD was taken at 260 nm and 280 nm to determine sample concentration and purity. Use of degraded RNA can result in low yield, overrepresentation of the 5' ends of the RNA molecules, or failure of the whole transcriptome sequencing library preparation. Total RNA integrity was checked following isolation using an Agilent Technologies 2100 Bioanalyzer with an RNA Integrity Number (RIN) value greater than 7.0.

RNA was quantified with rRNA ratio≧1.5, amount of RNA≧1.2ug and RIN>5.0. Ideal RIN of most RNA samples was >7.0, but if repeated samples cannot reach 7.0 of RIN, 5.0 of RIN was used as marginal cut-off. Extraction of high-quality RNA and preparation of library was rigorously repeated until every RNA sample meets all above quality criteria as the starting material.

4. Whole transcriptome sequencing of SNU cohort

For SNU cohort, all tumor samples from SNU cohort were prepared to whole transcriptome library using Illumina Truseq RNA library preparation kit (Ribo-Zero rRNA Removal Kit). All libraries were sequenced on Illumina HiSeq2000 platform using one sample per lane, with a paired-end 2 x 101 bp read length. Tumor RNA and its corresponding normal RNA were usually loaded on the same flow cell. At least 10 gigabytes of RAW data per each sample were generated and were converted to the FASTQ format. Read alignment and processing were performed using STAR aligner and Picard at the Broad Institute (http://broadinstitute.github.io/picard/) as GATK best practice recommendation (45). Expression of mRNA was quantified using de-duplicated BAM files by FPKM (fragments per kilobase of exon per million mapped reads) using HTSeq-count based on the Homo Sapiens GRCh37 Ensemble v65 (46).

5. Whole exome sequencing of SNU cohort

For SNU cohort, whole-exome sequencing of at least 3 ug of dsDNA from tumor and its corresponding normal gastric mucosa samples was performed using Agilent SureSelect Human All Exon V5 + UTR region kit. A paired-end 2 x 101 bp reads were sequenced on Illumina HiSeq2000 platform. On target depth of sequencing was planned as at least 100x for both tumors and normal mucosa (ideally 200x for tumors) . Read alignment and processing were performed using the Burrows-Wheeler Aligner (BWA)-mem and Picard at the Broad Institute as GATK best practice recommendation (47, 48).

6. Predictive classification algorithms using transcriptome sequencing

We used BRB Array Tools for analysis gene expression data(49). RNA-sequencing data from TCGA and SNU cohorts were analyzed together to identify differentially expressed genes (DEG) and construct prediction model. Firstly, DEGs between EAC and GCFB in TCGA cohort were identified by Student's *t* test (P<0.001), and further selected according to top and bottom fold change rank. For construction of prediction model, we used previously established Bayesian Compound Covariate Predictor (BCCP) algorithm with use of the leave-one-out cross validation (LOOCV) approach (50, 51) (Figure 3).

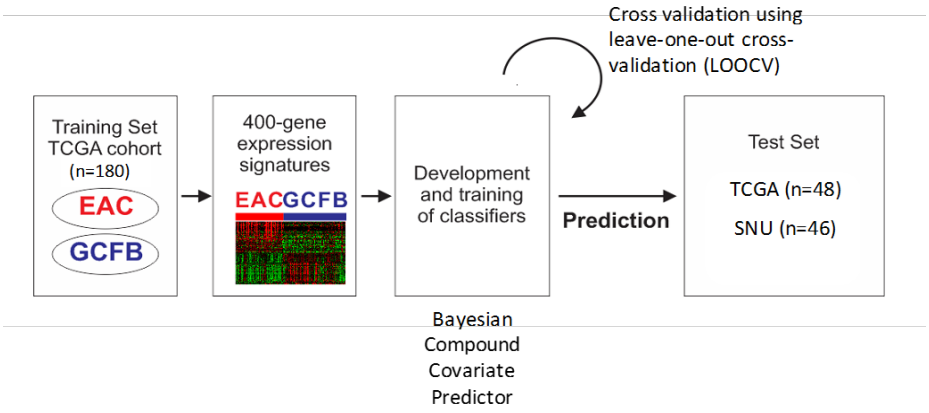


Figure 3. Class prediction model with Bayesian compound covariate predictor.

Sensitivity and specificity of trained model was evaluated by the Receiver Operating Characteristics (ROC) curve. Optimal cut-off value between EAC and GCFB was determined using Youden index. External validation for prediction model with cut-off value was performed using independent RNA microarray data of gastric and esophageal adenocarcinoma cell lines from Cancer Cell Line Encyclopedia (CCLE) database (http://www.broadinstitute.org/ccle). According to likelihood of BCCP model, GEJ/cardia tumors of TCGA cohort and all tumors of SNU cohort were reclassified in genomic subtypes (EAC-like or GCFB-like groups). Difference between genomic subtypes in clinical and molecular level were later assessed by analyzing clinicopathologic data, mutations, copy number alteration. Pathway analysis was carried out by using the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis tool (http://www.kegg.jp/)(52). Potential surrogate markers associated with genomic subtypes were validated by reverse-phase protein assay for

TCGA cohort and tissue microarray for SNU cohort. Target drug responsiveness of those surrogate markers were compared using IC50 of gastric and esophageal adenocarcinoma cell lines from CCLE database.

7. Identification of somatic mutations and insertion/deletions

For TCGA cohort, somatic mutation including insertion/deletions were analyzed using previously reported method (41, 42).

For SNU cohort, the BAM files for whole exome sequencing were used for somatic mutation calling using Mutect and IndelGenotyper (53). Variants with 1) exonic and splicing variants based on the reference sequence or variants with 2) more than 8 read depths and more than 4 alternate allele depths were selected. Variants with common variants of dbSNP142 or with population frequencies of > 0.01 in The Exome Aggregation Consortium, 1000 Genomes Project and NHLBI ESP6500 were excluded (54-56). Functional annotation of mutations was performed with ANNOVAR. Significantly recurrently mutated genes were identified using the MutSigCV2.0 algorithm (57). We compared somatic mutation and insertion/deletions between EAC-like and GCFB-like subgroup in each TCGA and SNU cohort.

8. Somatic copy number analysis

For TCGA cohort, copy-number alterations (CNAs) data from single-nucleotide polymorphism (SNP) array were analyzed using previously

reported method (41, 42). For SNU cohort, CNAs were analyzed using whole exome data based on the RPKM (Read Per Kilobase per Million mapped reads) value from CONIFER (58). Analysis of somatic CNAs was performed with the GISTIC 2.0 algorithm for both TCGA and SNU cohort. Among genes with focal copy number amplification using GISTIC algorithm, we selected candidate genes with log2 copy number ratio of tumor over corresponding normal gastric mucosa $\geqq 1$ in at least one paired sample. We compared copy number of those candidate genes between EAC-like and GCFB-like subgroup using Student's *t* test.

9. Reverse-phase protein array of TCGA cohort

Reverse-phase protein array (RPPA) data of 132 out of 180 cases (comprised of 44 EAC and 88 GCFB) were retrieved in database of the Cancer Genome Atlas (TCGA). Clustering analysis was performed after re-centered normalization.

10. Tissue microarray of SNU cohort

Tissue microarray (TMA) was assembled according to the following procedure: Core tissue biopsies (diameter 2 mm) were obtained from individual paraffin-embedded gastric tumors (donor blocks) and arranged in new recipient paraffin blocks (tissue array blocks) using a trephine apparatus (Superbiochips Laboratories, Seoul, Korea). Considering the possible diversity of histological components or molecular abnormality of advanced cancer, we developed 3 sets of TMA

for each sample. The tissue array blocks contained up to 46 cores on 3 arrays, for a total of 138 cases for immunohistochemistry (IHC) staining. Tumors occupying more than 10% of the core area were considered adequate. Each paraffin block contained internal controls, which consisted of non-neoplastic gastric mucosa from the body and antrum as well as intestinal metaplasia. IHC was performed using an automatic immunostainer (BenchMark XT, Ventana Medical Systems, Tucson, AZ, USA), as described by the manufacturer's protocol. After tissues were sampled from in each core, staining intensity were scored as 0 (no membrane staining, negative), 1+ (faint/barely perceptible partial staining, weakly positive), 2+ (weak-to-moderate staining, moderately positive) and 3+ (Strong staining, strongly positive). All immunohistochemical staining and silver in situ hybridization (SISH) for each TMA core was assessed and scored by one expert pathologist unaware of any clinical information. Staining status for all proteins except for ERBB2 were analyzed using complex H-score by multiplying the staining intensity by the percentage of cells stained and the sum of individual H-scores for each intensity level (59). To decrease possible tumor heterogeneity inside each tumor, 3 TMA cores from each sample were regarded as tentatively different samples, and average complex H-score between triplicated EAC-like and GCFB-like groups was compared as using Student's *t* test. Staining status of ERBB2 was regarded as positive as the highest stain intensity score when $\geqq$ 10% of cells were stained as that intensity in at least one TMA core. Final

interpretation using the results of IHC and SISH was conducted as previously reported method (60-62). Expression of ERBB2 was dichotomized into positive as IHC 3+ or IHC2+ and black/red ratio of SISH ≧2.0, and negative as IHC<2+, or IHC 2+ and black/red ratio of SISH <2.0. Target drug response was evaluated using IC50 data of CCLE database for gastric and esophageal adenocarcinoma cell lines.

11. Statistical analysis

Student's *t* test and a chi-squared test were used for comparative analysis. Survival analysis was conducted using the Kaplan-Meier method and the log rank test. Multivariate analysis to identify risk factors for protein expression was conducted using binary logistic regression or linear regression analysis with method of backward. All tests were 2-sided and performed at a significance level of 5% using SPSS version 21.0 (SPSS, Inc., Chicago, IL, USA).

RESULTS

Figure 4 demonstrated detailed study population according to analysis scheme. We analyzed 228 tumors of pure esophageal adenocarcinoma, pure gastric adenocarcinoma at fundus/body, adenocarcinoma at GEJ/cardia from TCGA cohort, and 46-paired (92 samples) tumors-corresponding normal mucosa of AGEJ II, AGEJ III, and UT from SNU cohort.

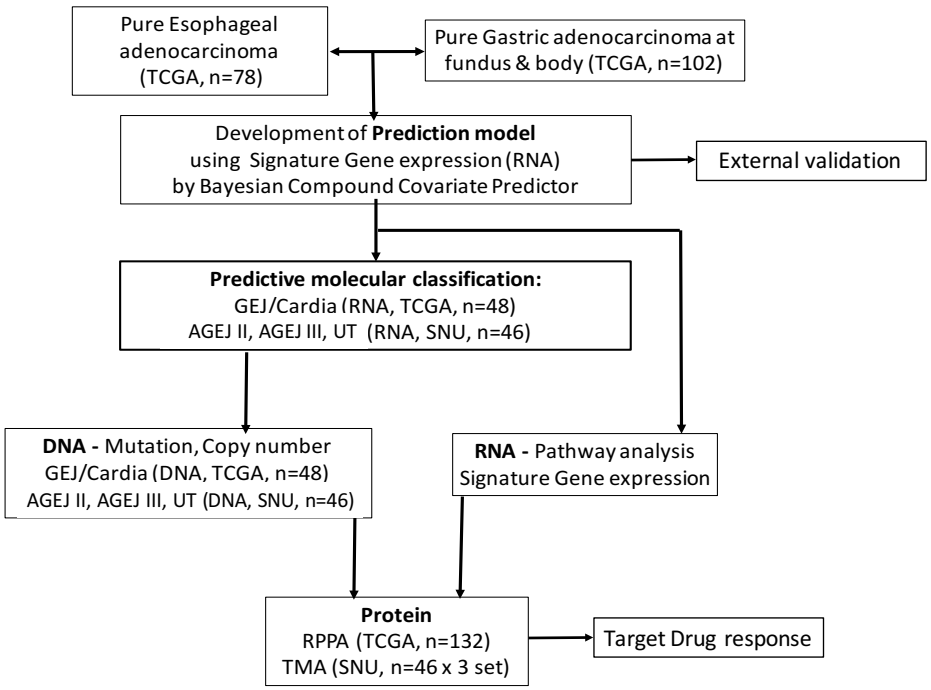


Figure 4. Detailed study population according to analysis scheme.

For SNU cohort, after repeated extraction of nucleic acid from fresh frozen tissue and library preparation with meticulous quality control, we successfully retrieved raw sequencing data from all planned samples except for 2 UT samples (Table 1).

Table 1. Quality of sequencing data for whole transcriptome and exome.

		Transcriptome					Exome	
	serial	Total number of bases sequenced	Total number of reads sequenced	GC content (%)	Ratio of reads that have phred quality score of over 20 (%)	Ratio of reads that have phred quality score of over 30 (%)	Mean coverage	Coverage >100x (%)
AGEJII	AGEJII1N	10,716,808,616	106,107,016	53.258	96.447	93.157	112.08	50.7
	AGEJII1T	11,155,658,666	110,452,066	48.719	96.508	93.269	107.3	48.3
	AGEJII2N	10,854,630,388	107,471,588	52.094	96.44	93.172	106.21	47.4
	AGEJII2T	11,428,892,956	113,157,356	48.167	96.188	92.944	101.37	44.8
	AGEJII3N	12,032,274,834	119,131,434	52.634	96.008	92.474	101.98	44.7
	AGEJII3T	11,887,353,772	117,696,572	51.953	96.663	93.454	141.61	65.2

AGEJII4N	11,521,509,552	114,074,352	50.383	96.362	93.066	113.78	50.1
AGEJII4T	11,264,305,780	111,527,780	49.08	94.78	90.40	83.1	28.7
AGEJII5N	10,117,392,806	100,172,206	50.21	95.16	91.10	113.65	51.8
AGEJII5T	12,638,081,924	125,129,524	50.983	95.55	91.711	115.93	52.1
AGEJII6N	12,148,675,516	120,283,916	51.155	96.449	93.15	112.49	52.2
AGEJII6T	12,371,802,292	122,493,092	49.601	96.568	93.411	114.03	52.2
AGEJII7N	10,673,791,302	105,681,102	53.027	96.772	93.693	53.11	13.2
AGEJII7T	10,463,644,238	103,600,438	52.364	96.852	93.83	67.21	21
AGEJII8N	10,049,880,568	99,503,768	51.85	96.097	92.547	127.37	52.2
AGEJII8T	12,244,691,772	121,234,572	50.141	96.404	93.1	240.75	79.3

AGEJII9N	12,074,532,830	119,549,830	51.842	96.392	93.057	135.26	53.1
AGEJII9T	10,268,023,398	101,663,598	48.74	95.35	91.29	236.58	69.7
AGEJII10N	11,182,880,388	110,721,588	52.539	96.411	93.078	129.91	48.2
AGEJII10T	11,824,151,204	117,070,804	50.592	96.146	92.921	214.39	83.5
AGEJII11N	10,152,334,766	100,518,166	53.99	95.901	95.901	131.36	53.4
AGEJII11T	11,316,389,662	112,043,462	49.541	96.408	93.122	269.2	82.2
AGEJII12N	11,020,593,184	109,114,784	49.83	95.50	91.74	130.61	54.3
AGEJII12T	10,036,132,246	99,367,646	49.41	95.44	91.59	267.87	83.3
AGEJII13N	10,445,388,690	103,419,690	51.84	95.33	91.17	114.12	47.2
AGEJII13T	10,084,313,690	99,844,690	51.46	95.18	90.95	248.06	84

	AGEJII14N	10,167,142,578	100,664,778	51.24	95.06	91.45	126.17	51.9
	AGEJII14T	10,806,255,226	106,992,626	49.03	95.58	91.70	244.06	77.7
	AGEJII15N	10,459,037,426	103,554,826	50.37	95.38	91.32	116.37	48.1
	AGEJII15T	11,730,982,138	116,148,338	50.27	95.40	91.67	287.81	82
	AGEJII16N	10,714,711,048	106,086,248	51.45	95.03	90.76	143.14	56.2
	AGEJII16T	10,286,269,856	101,844,256	49.0	95.19	91.15	258.48	82.9
AGEJ III	AGEJIII01N	14,452,581,668	143,094,868	49.55	95.93	91.90	85.18	33
	AGEJIII01T	15,890,118,304	157,327,904	49.36	96.04	92.18	110.01	45.8
	AGEJIII02N	13,656,118,292	135,209,092	51.08	95.94	91.93	78.11	27.7

AGEJIII02T	14,656,213,424	145,111,024	49.58	96.10	92.25	74.67	25.4
AGEJIII03N	13,705,105,716	135,694,116	50.47	95.52	91.18	72.2	24.4
AGEJIII03T	14,652,128,580	145,070,580	50.10	97.15	95.19	39.91	3.2
AGEJIII04N	11,978,208,524	118,596,124	51.50	95.73	91.54	131.65	59
AGEJIII04T	21,175,946,032	209,662,832	50.39	96.94	94.82	110.98	47.2
AGEJIII05N	11,861,906,418	117,444,618	49.90	95.87	91.84	128.65	51.7
AGEJIII05T	12,628,785,278	125,037,478	49.69	95.97	92.02	267.31	80.2
AGEJIII06N	12,637,132,322	125,120,122	52.11	96.02	92.10	156.43	58.6
AGEJIII06T	13,409,600,522	132,768,322	49.27	96.01	92.16	243.23	80.4
AGEJIII07N	13,727,712,748	135,917,948	49.35	95.98	92.11	144	55.8

AGEJIII07T	13,977,555,640	138,391,640	50.04	96.98	94.91	236.65	76.7
AGEJIII08N	13,147,619,652	130,174,452	52.41	97.08	95.05	128.87	51.9
AGEJIII08T	12,446,576,632	123,233,432	51.33	95.76	91.66	250.74	80.1
AGEJIII09N	13,540,072,524	134,060,124	50.22	96.0	92.15	143.6	54.8
AGEJIII09T	15,704,037,520	155,485,520	50.21	96.98	94.91	217.69	79.1
AGEJIII10N	12,636,106,364	125,109,964	50.77	95.97	92.04	145.65	54.5
AGEJIII10T	12,706,649,612	125,808,412	50.46	96.0	92.12	232.68	80.5
AGEJIII11N	13,395,876,440	132,632,440	52.67	95.69	91.44	140.36	55.3
AGEJIII11T	13,446,510,972	133,133,772	50.71	97.02	94.97	241.49	80.6
AGEJIII12N	11,766,698,768	116,501,968	49.66	96.65	94.36	119.96	46.5

	AGEJIII12T	16,130,073,700	159,703,700	50.20	96.93	94.82	224.48	78.6
	AGEJIII13N	12,861,840,960	127,344,960	49.77	95.90	91.93	144.11	53.3
	AGEJIII13T	12,400,576,788	122,777,988	50.49	95.96	92.04	251.97	79.4
	AGEJIII14N	13,221,249,662	130,903,462	51.32	95.78	91.68	120.71	47.1
	AGEJIII14T	14,678,176,682	145,328,482	49.36	96.05	92.19	235.55	73.8
	AGEJIII15N	12,190,985,224	120,702,824	51.03	95.49	91.15	133.64	55
	AGEJIII15T	13,726,808,798	135,908,998	50.44	95.86	91.84	225.27	75.5
	AGEJIII16N	12,607,774,854	124,829,454	50.08	95.98	92.06	126.16	52.6
	AGEJIII16T	12,653,613,502	125,283,302	50.72	95.67	91.50	238.89	76.3
UT	UT1N	10,736,253,540	106,299,540	49.59	94.55	89.86	47.05	6.7

UT1T	12,551,428,166	124,271,566	49.50	94.56	89.89	37.05	2.3
UT2N	12,221,573,680	121,005,680	51.39	94.17	89.14	114.49	52
UT2T	11,964,286,078	118,458,278	50.91	94.28	89.35	93.41	39.7
UT3N	14,273,222,232	141,319,032	50.38	95.97	92.03	73.73	25.2
UT3T	11,586,835,342	114,721,142	50.01	94.58	89.87	37.12	3.8
UT4N	13,119,098,060	129,892,060	50.27	94.51	89.79	40.58	3.4
UT4T	12,372,752,500	122,502,500	50.54	94.57	89.91	36.4	2.1
UT5N	13,125,537,214	129,955,814	50.23	94.40	89.65	122.08	54.7
UT5T	12,061,402,628	119,419,828	51.23	94.49	89.79	113.67	48.5
UT6N	12,930,285,630	128,022,630	51.53	94.11	89.05	100.02	41.3

UT6T	11,749,007,002	116,326,802	50.42	94.52	89.76	78.62	24.7
UT7N	13,125,709,318	129,957,518	50.12	94.55	89.89	90.66	35.9
UT7T	11,994,315,600	118,755,600	51.25	94.37	89.56	191.73	70.3
UT9N	11,669,452,736	115,539,136	49.28	96.26	91.56	92.1	39.1
UT9T	11,533,054,256	114,188,656	48.07	96.48	92.07	90.2	36.9
UT10N	12,370,758,962	122,482,762	51.91	94.77	90.13	117.09	49.9
UT10T	13,422,684,264	132,897,864	48.92	95.92	91.96	189.53	75.8
UT11N	14,929,673,752	147,818,552	52.30	95.94	91.90	125.2	55
UT11T	13,875,930,854	137,385,454	49.67	95.90	91.92	212.5	80.1
UT12N	14,259,757,316	141,185,716	49.40	95.95	91.99	136.87	60.7

UT12T	14,852,647,314	147,055,914	52.32	97.07	95.05	192.81	76.4
UT13N	12,287,157,424	121,655,024	49.33	95.95	92.02	129.41	57
UT13T	12,502,802,120	123,790,120	50.11	96.02	92.11	196.09	78.4
UT14N	12,132,667,420	120,125,420	51.61	95.84	91.73	122.64	53
UT14T	12,102,652,846	119,828,246	51.07	95.89	91.85	205.25	81.1
UT15N	12,371,734,218	122,492,418	50.90	95.59	91.24	130.24	56.9
UT15T	12,283,873,712	121,622,512	50.79	95.51	91.10	186.12	75.9

Clinicopathologic characteristics

In TCGA cohort, we identified 78 EAC, 48 GEJ/Cardia and 102 GCFB samples available for exome and transcriptome data (Table 2).

Table 2. Clinicopathologic characteristics of TCGA cohort.

		EAC (n=78)	GEJ/Cardia (n=48)	GCFB (n=102)	*P* value
Gender (Male : Female)		69:9	37:11	57:45	<0.001
Age (years)		66.8±12.0	66.9±9.2	66.6±9.3	0.985
Location	Esophagus, mid	2 (2.6%)	0	0	<0.001
	Esophagus, mid-distal	1 (1.3%)	0	0	
	Esophagus, distal	75 (96.2%)	0	0	
	GEJ/cardia	0	48 (100%)	0	
	Fundus/Body	0	0	102 (100%)	
WHO	Papillary	0	4 (8.3%)	12 (11.8%)	
	Tubular	0	23 (47.9%)	49 (48.0%)	
	Poorly Cohesive	0	9 (18.8%)	19 (18.6%)	
	Mucinous	0	3 (6.3%)	4 (3.9%)	
	Mixed	0	6 (12.5%)	6 (5.9%)	
	not available	78 (100%)	3 (6.3%)	12 (11.8%)	
Lauren	Intestinal	0	32 (66.7%)	70 (68.6%)	

	Diffuse	0	9 (18.8%)	19 (18.6%)	
	Mixed	0	6 (12.5%)	6 (5.9%)	
	not available	78 (100%)	1 (2.1%)	7 (6.9%)	
T stage	T1	20 (25.6%)	1 (2.1%)	7 (6.9%)	
	T2	10 (12.8%)	18 (37.5%)	17 (16.7%)	
	T3	34 (43.6%)	24 (50.0%)	53 (52.0%)	
	T4	0	2 (4.2%)	0	
	T4a	0	1 (2.1%)	19 (18.6%)	
	T4b	0	1 (2.1%)	6 (5.9%)	
	TX	14 (17.9%)	1 (2.1%)	0	
N stage	N0	19 (24.4%)	15 (31.3%)	41 (40.2%)	
	N1	36 (46.2%)	16 (33.3%)	17 (16.7%)	
	N2	5 (6.4%)	6 (12.5%)	14 (13.7%)	
	N3	4 (5.1%)	9 (18.8%)	25 (24.5%)	
	NX	14 (17.9%)	2 (4.2%)	5 (4.9%)	
M stage	M0	44 (56.4%)	41 (85.4%)	95 (93.1%)	
	M1	5 (6.4%)	3 (6.3%)	5 (4.9%)	
	MX	29 (37.2%)	4 (8.3%)	2 (2.0%)	
Country	Australia	1 (1.3%)	0	0	<0.001
	Brazil	2 (2.6%)	0	0	
	Canada	8 (10.3%)	2 (4.2%)	0	
	Germany	0	8 (16.7%)	12 (11.8%)	
	Korea South	0	1 (2.1%)	8 (7.8%)	
	Netherlands	9 (11.5%)	0	0	
	Poland	0	6 (12.5%)	14 (13.7%)	
	Russia	0	7 (14.6%)	49 (48.0%)	

Ukraine	1 (1.3%)	9 (18.8%)	11 (10.8%)
United Kingdom	1 (1.3%)	0	0
United States	56 (71.8%)	11 (22.9%)	3 (2.9%)
Vietnam	0	4 (8.3%)	5 (4.9%)

Race of samples were significantly different among each 3 group. The proportion of East Asian countries including Korea and Vietnam in overall samples was 18/228 (7.9%), and that in GEJ/Cardia was 5/48 (10.4%). There was no sample from East Asian countries in EAC group. In SNU cohort, we collected 16 AGEJ II, 16 AGEJ III and 14 UT tumor samples and its corresponding normal gastric mucosa (Table 3).

Table 3. Clinicopathologic characteristics of SNU cohort.

		AGEJ II (n=16)	AGEJ III (n=16)	UT (n=14)	*P* value
Gender (M:F)		13:3	12:4	11:3	0.912
Age (years)		58.5±10.4	66.5±9.4	63.5±8.1	0.062
WHO classification	Differentiated	7 (43.8%)	7 (43.8%)	7 (50.0%)	0.919
	Undifferentiated	7 (43.8%)	8 (50.0%)	5 (35.7%)	

	Undetermined	2 (12.5%)	1 (6.3%)	2 (14.3%)	
Lauren classification	Intestinal	5 (31.3%)	6 (37.5%)	7 (50.0%)	0.526
	Diffuse	7 (43.8%)	5 (31.3%)	6 (41.9%)	
	Mixed	4 (25.0%)	5 (31.3%)	1 (7.1%)	
Lymphatic invasion		12 (75.0%)	9 (56.3%)	9 (64.3%)	0.600
Venous invasion		4 (25.0%)	3 (18.8%)	4 (28.6%)	0.873
Perineural invasion		10 (62.5%)	13 (81.3%)	8 (57.1%)	0.326
Tumor size (cm)		4.9±1.5	7.6±3.8	6.4±2.9	0.035
T stage	T1	1 (6.3%)	0	0	0.311
	T2	3 (18.8%)	1 (6.3%)	4 (33.3%)	
	T3	8 (50.0%)	6 (37.5%)	4 (33.3%)	
	T4a	4 (25.0%)	7 (43.8%)	4 (33.3%)	
	T4b	0	2 (12.5%)	0	

N stage	N0	2 (12.5%)	4 (25.0%)	4 (28.6%)	0.138
	N1	2 (12.5%)	0	4 (28.6%)	
	N2	3 (18.8%)	5 (31.3%)	4 (28.6%)	
	N3	9 (56.3%)	7 (43.8%)	2 (14.2%)	
M stage	M0	15 (93.8%)	14 (87.5%)	13 (92.9%)	0.390
	M1	1 (6.3%)	2 (12.5%)	1 (7.1%)	
TNM stage	I	1 (6.3%)	1 (6.3%)	2 (14.3%)	0.549
	II	3 (18.8%)	3 (18.8%)	5 (35.7%)	
	III	11 (68.8%)	10 (62.5%)	6 (42.9%)	
	IV	1 (6.3%)	2 (12.5%)	1 (7.1%)	
Neoadjuvant chemotherapy		1 (6.3%)	1 (6.3%)	0	0.633
Microsatellite instability	MSS	12 (75.0%)*	12 (75.0%)	11 (78.6%)	0.381
	MSI-Low	2 (12.5%)	3 (18.8%)	0	
	MSI-high	0	0	2 (14.3%)	

Not available	2 (12.5%)	1 (6.3%)	1 (7.1%)

* 2 patients; only BAT26 (-) available

Race of all samples was Asian (Korean). Proportion of differentiation or Lauren classification was not significantly different among each 3 group. Average tumor size of AGEJ III is significantly larger than that of AGEJ II (*P*=0.014), but not different from that of UT (*P*=0.326). Regarding proportion of stage, stage I is 6.3% for AGEJ II or AGEJ III, and 14.3% for UT. There was no MSI-high in AGEJ II or AGEJ III.

Clustering analysis of SNU cohort based on anatomic subgroup

Unsupervised clustering of whole transcriptome data of SNU cohort showed clear separation between tumor and normal, but no distinctive separation pattern according to anatomic subgroup (Figure 5).

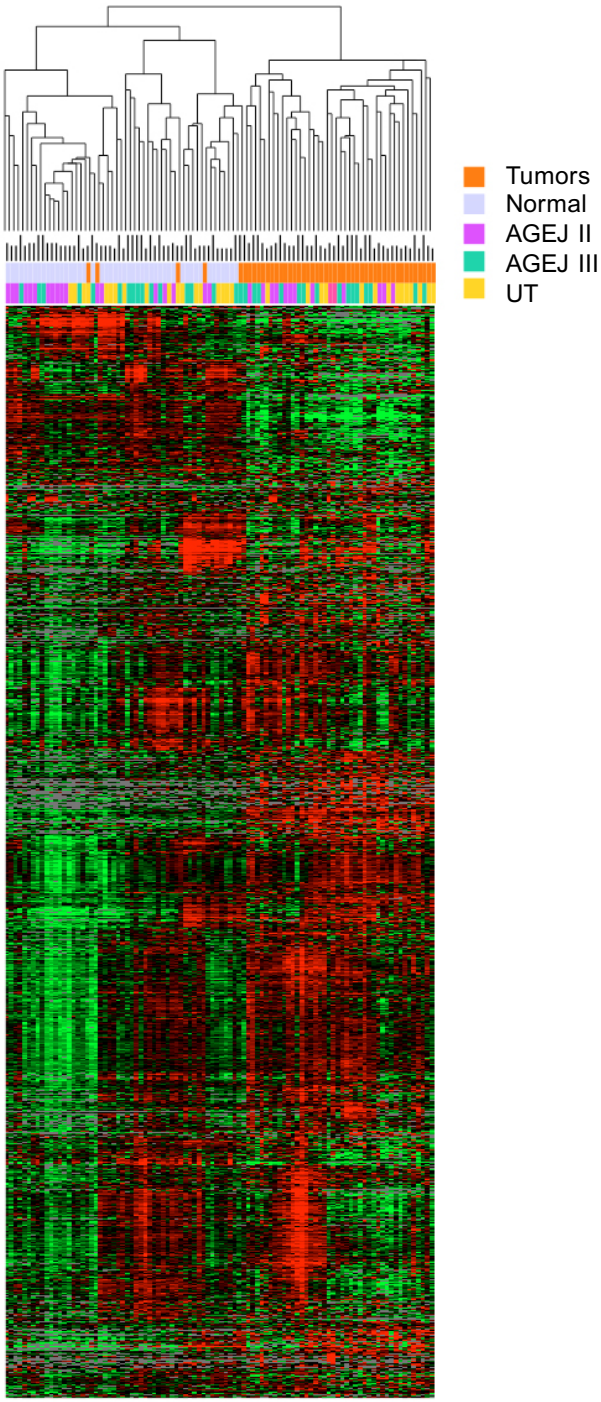


Figure 5. Unsupervised hierarchical clustering of AGEJ II, AGEJ III, and UT in SNU cohort between tumor and normal samples

When we clustered tumors only of SNU cohort, two molecular subgroups were clustered but failed to show any significant separation based on anatomic subgroups (Figure 6).

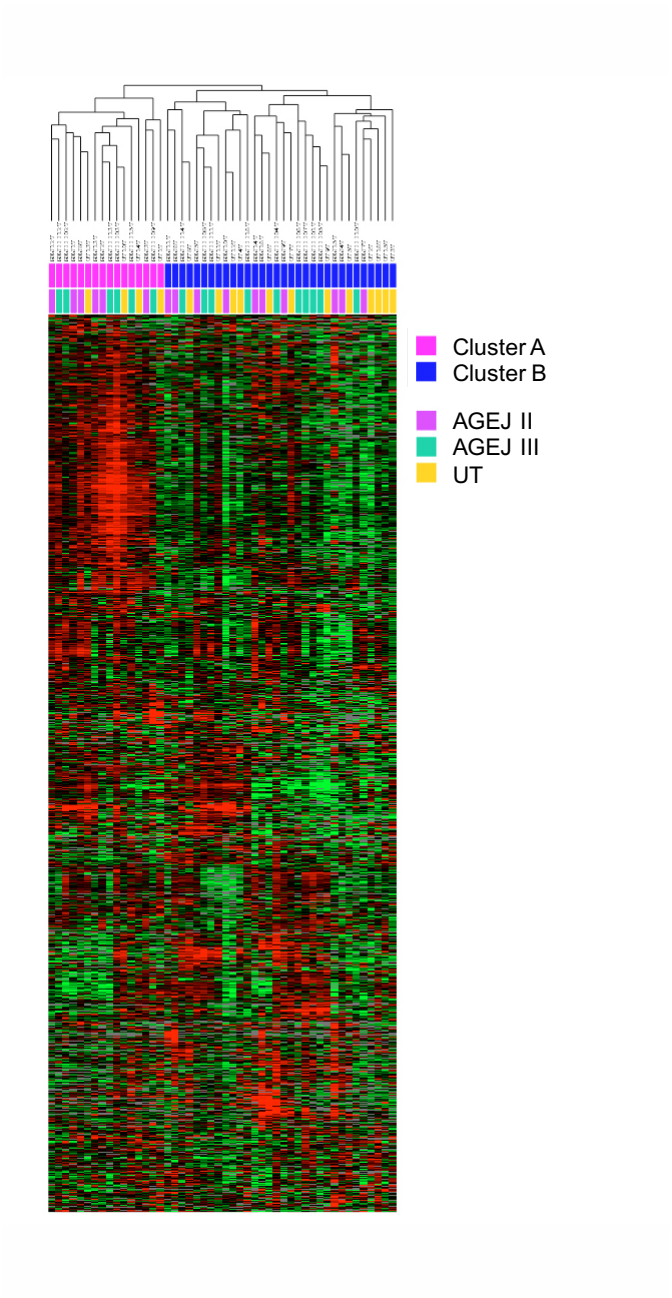


Figure 6. Unsupervised hierarchical clustering of tumors only in AGEJ II, AGEJ III, and UT in SNU cohort.

When previous 4 molecular subgroups of TCGA were applied for clustering, there was no definitive correlation according to anatomic subgroups(41) (Figure 7).

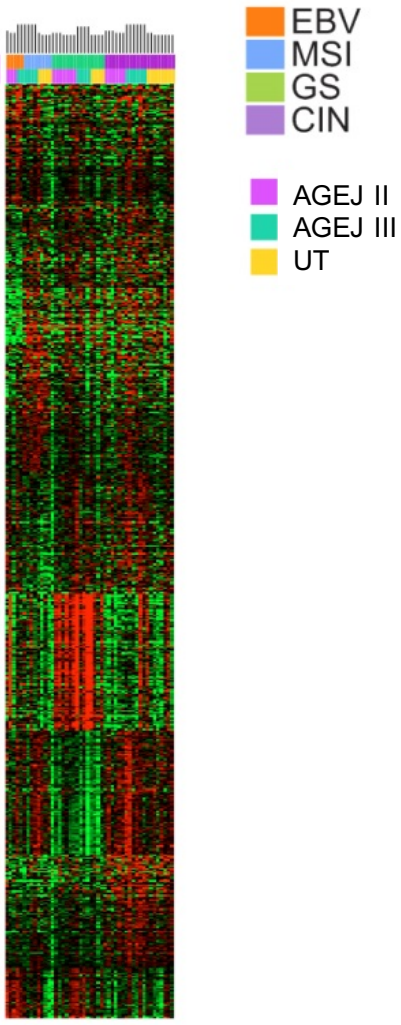


Figure 7. Unsupervised hierarchical clustering of tumors in AGEJ II, AGEJ III, and UT in SNU cohort according to TCGA 4 subgroups.

Development of predictive classification model

Unsupervised hierarchical clustering of EAC and GCFB in TCGA cohort revealed 5,520 genes with *P*<0.001 (Figure 8).

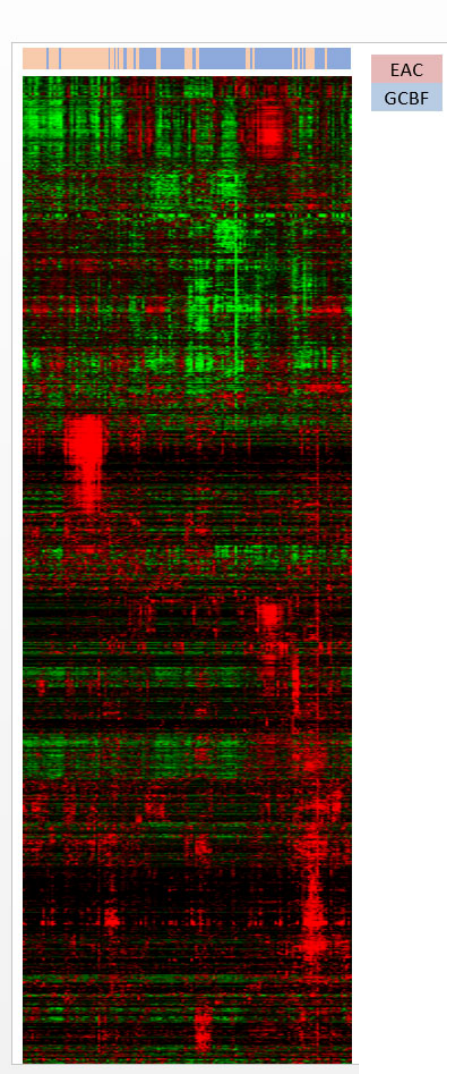


Figure 8. Unsupervised clustering with 5,520 genes between esophageal adenocarcinoma and gastric cancer at fundus or body in TCGA cohort.

According to fold change rank, each top 200 and bottom 200 genes were selected as 400 signature gene classifiers. We performed unsupervised hierarchical clustering of EAC and GCFB in TCGA cohort using these 400 signature gene classifiers and identified clear separation of clusters between EAC and GCFB (Figure 9).

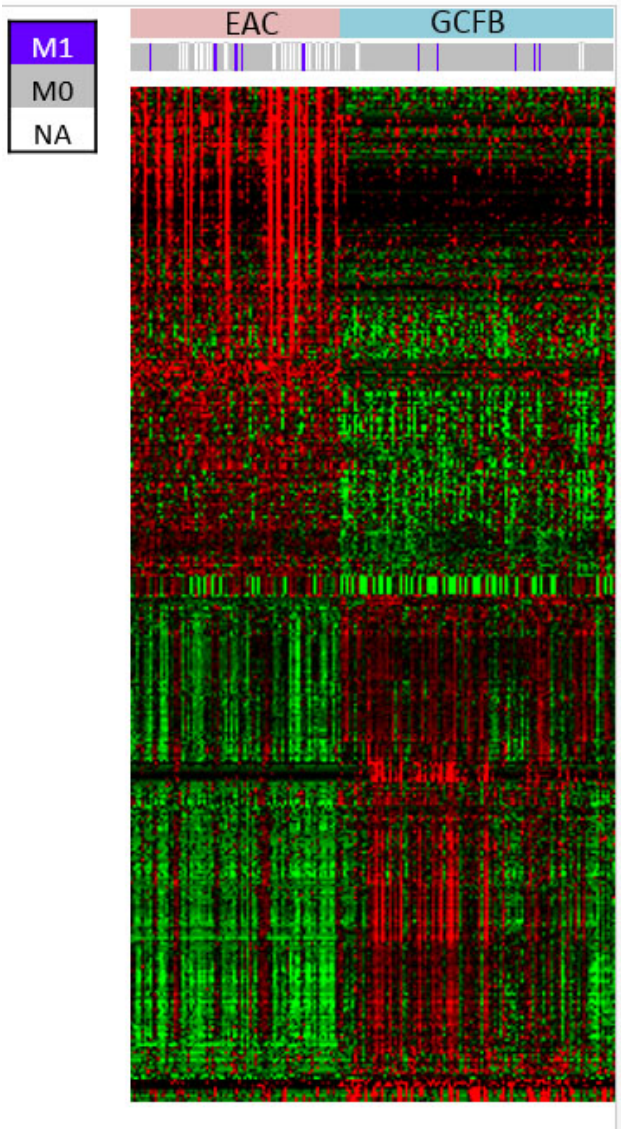


Figure 9. Heatmap between esophageal adenocarcinoma and gastric

cancer at fundus or body from TCGA training cohort using 400 signature classifier genes.

Predictive classification model was developed based on BCCP with 400 signature gene classifiers and trained by LOOCV. ROC curve using BCCP scores revealed 0.957 of area under curve (95% confidence interval=0.93-0.98), and 0.4535 of Youden index as a cut-off value between EAC and GCFB (Figure 10).

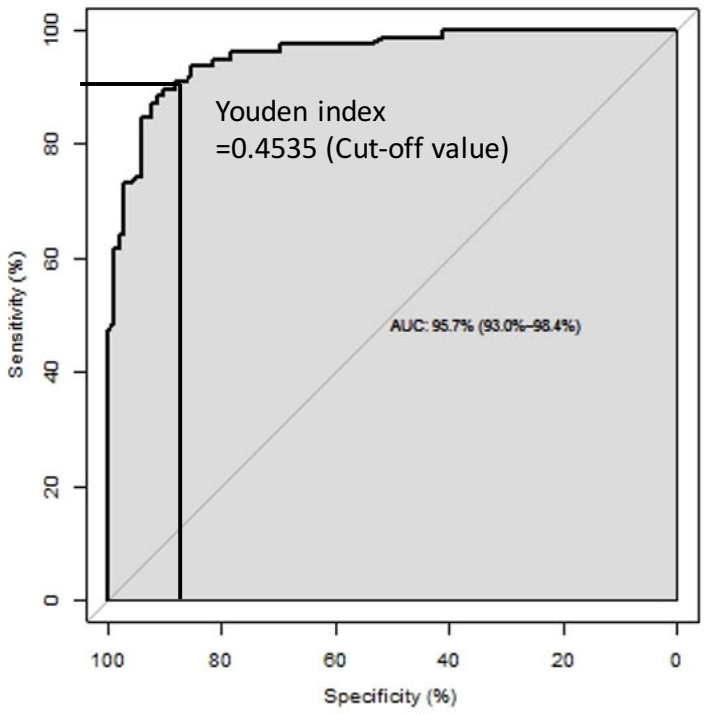


Figure 10. ROC curve after cross validation using Leave-one-out cross validation.

That cut-off value demonstrated 90.2% of sensitivity and 89.7% of specificity to predict EAC. For those 400 signature genes, pathway analysis was conducted using KEGG pathway analysis. Among several cancer-related pathways with 5 or more genes involved, we identified PI3K-AKT signaling pathway related to GCFB in which CHRM2, COMP, FGF14, IGF1, PPP2R2B, RELN, THBS4 out of overexpressed 200 genes for GCFB were involved. Consequently, PI3K and AKT were considered for protein validation using RPPA of TCGA cohort and tissue microarray of SNU cohort.

Test of predictive classification model with somatic mutation analysis

Using BCCP scores with 0.4535 as a cut-off value, we tested clustering for GEJ/cardia of TCGA cohort (Figure 11).

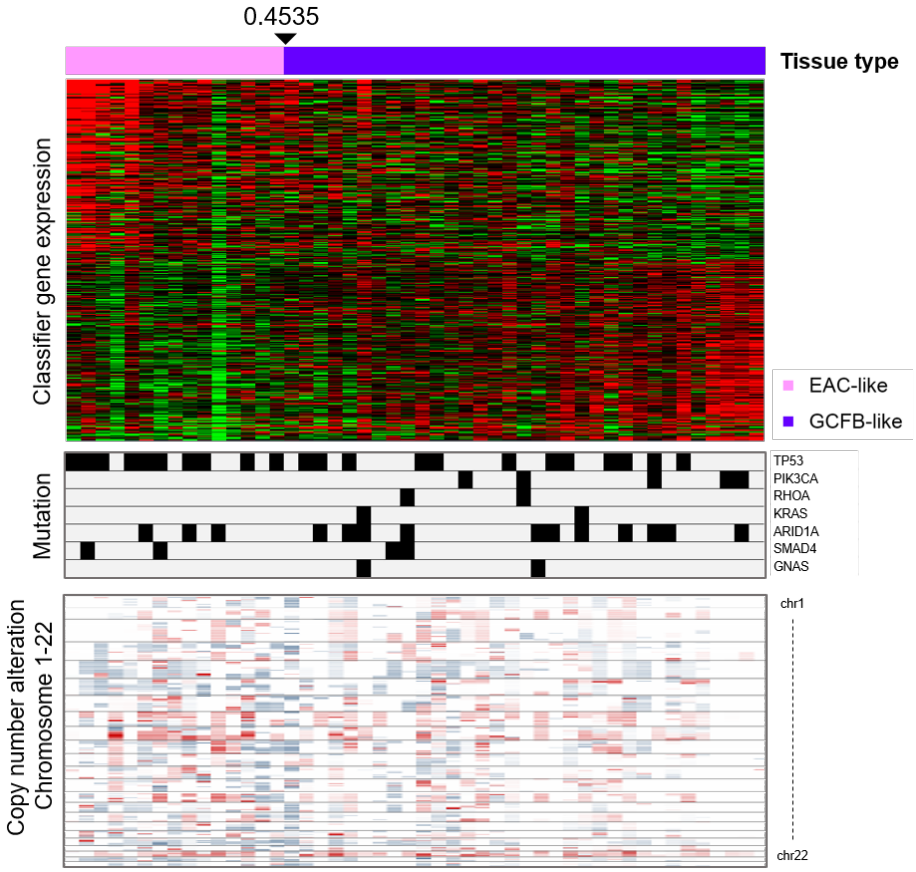


Figure 11. Hierarchical clustering of GEJ/Cardia in TCGA cohort using Bayesian compound covariate predictor.

Hierarchical clustering of GEJ/cardia of TCGA cohort shows spectral transition of clusters between EAC-like and GCFB-like group without any entirely distinguishable cluster. GEJ/cardia of TCGA cohort predicted as EAC was 15/48 (31.2%) and that predicted as GCFB was 33/48 (68.8%). In terms of somatic mutation, there was no significant difference of TP 53, PIK3CA, RHOA, KRAS, and ARID1A between EAC-like and GCFB-like group. When we tested clustering for AGEJ II, AGEJ III, UT of SNU cohort, SNU cohort also demonstrated similar spectral

transition of clusters between EAC-like and GCFB–like group, which is similar to TCGA cohort (Figure 12).

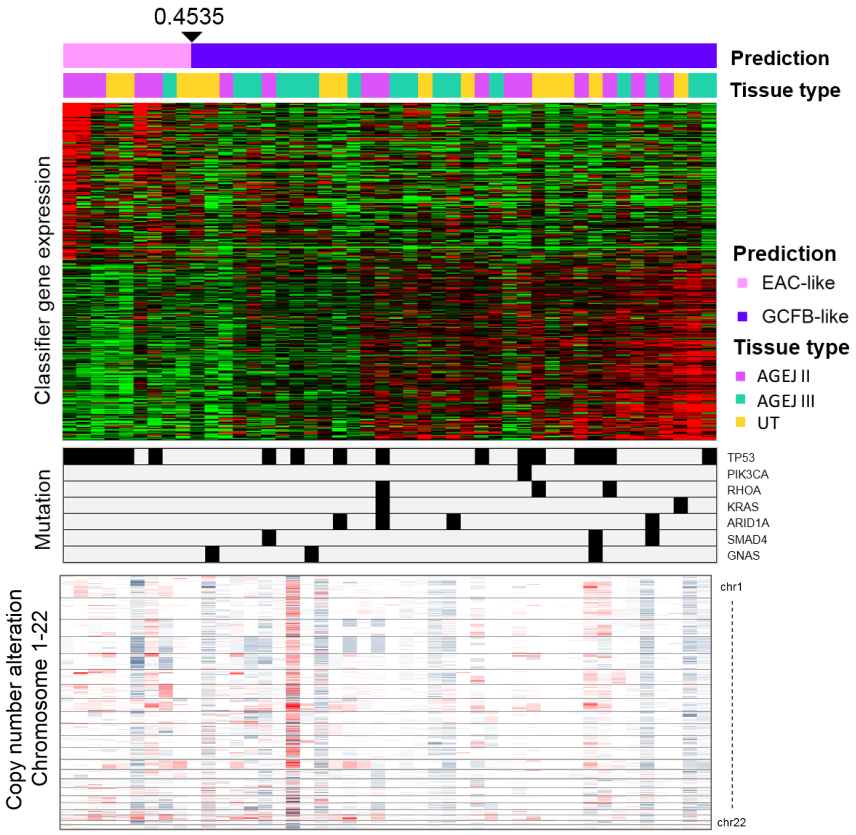


Figure 12. Hierarchical clustering of adenocarcinoma of gastroesophageal junction or upper third gastric cancer in SNU cohort Bayesian compound covariate predictor.

AGEJ II of SNU cohort was classified as 5/16 (31.2%) of EAC-like group and 11/16 (68.8%) of GCFB-like group. Especially, 15/16 (93.7%) of AGEJ III was classified as GCFB-like. Taken together with AGEJ II and III of SNU cohort, EAC-like and GCFB like was 6/32 (18.8%) and 26/32

(81.2%). There was also no significant difference of somatic mutation in genes including TP53, PIK3CA, ROHA, KRAS, ARID1A between EAC-like and GCFB-like in SNU cohort. Especially, any somatic mutation of RHOA, KRAS and PIK3CA was not found in EAC-group of both TCGA and SNU cohort.

Clinicopathologic analysis between EAC-like and GCFB-like group

Pathologic characteristics analysis of SNU cohort revealed that all AGEJ III involving GEJ and 80.0% (4/5) of AGEJ III without involving GEJ classified as GCFB-like group (Table 4).

Table 4. Pathologic characteristics between EAC-like and GCFB-like in SNU cohort.

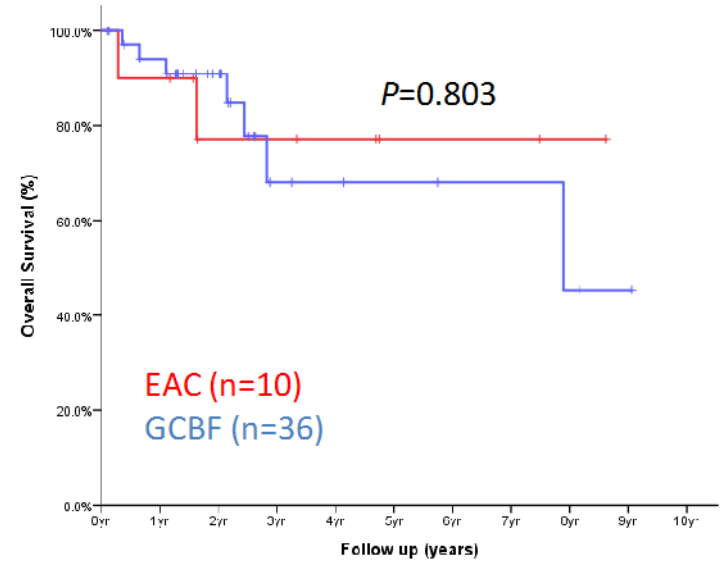
		EAC-like (n=10)	GCFB-like (n=36)	*P* value
Location	AGEJ II	5 (31.3%)	11 (68.8%)	0.231
	AGEJ III involving GEJ	0	11 (100%)	
	AGEJ III without involving GEJ	1 (20.0%)	4 (80.0%)	
	UT	4 (28.6%)	10 (71.4%)	

WHO	Differentiated	8 (80.0%)	13 (36.1%)	0.043
	Undifferentiated	2 (20.0%)	18 (50.0%)	
	undetermined	0	5 (13.9%)	
Lauren	Intestinal	8 (80.0%)	10 (27.8%)	0.009
	diffuse	2 (20.0%)	16 (44.4%)	
	mixed	0	10 (27.8%)	
Lymphatic invasion		4 (40.0%)	26 (72.2%)	0.107
Venous invasion		4 (40.0%)	7 (19.4%)	0.336
Perineural invasion		3 (30.0%)	28 (77.8%)	0.008
TNM	I	2 (20.0%)	2 (5.6%)	0.501
	II	3 (30.0%)	8 (22.2%)	

	II	5 (50.0%)	22 (61.1%)	
	IV	0	4 (11.1%)	
MSI	MSS	9 (90.0%)	24 (66.7%)	0.332
	MSI-L	0	4 (11.1%)	
	MSI-H	0	3 (8.3%)	
	N/A	1 (10.0%)	5 (13.9%)	

The distribution of EAC-like and GCFB-like was not significantly different among AGEJ II, AGEJ III and UT. However, EAC-like group shows significantly higher proportion of differentiated and intestinal type whereas GCFB-like group has significantly higher proportion of undifferentiated and diffuse type. There was no significant difference of TNM stage between EAC-like and GCFB-like groups. Postoperative overall survival as well as recurrence-free survival between both EAC-like and GCFB-like groups was not significantly different (Figure 13).

(A)



(B)

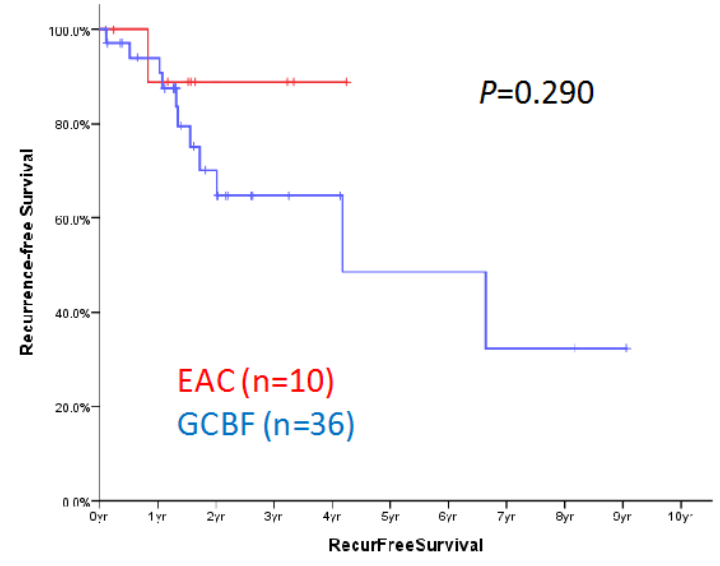


Figure 13. Postoperative survival between EAC-like and GCFB-like group in SNU cohort. (A) Overall survival in SNU cohort. (B) Recurrence-free survival in SNU cohort.

Copy number analysis between EAC-like and GCFC-like group

We performed genome-wide copy number analysis In TCGA cohort and

identified 435 amplified genes with significantly different copy number (≥ 2-fold change and *P*<0.05) between EAC-like and GCFB-like by the GISTIC algorithm. Filtration for those 435 genes by human Cancer Gene Census revealed 6 cancer-related genes including COX6C in 8q22.2 with translocation, HNRNPA2B1 in 7p15.2 with translocation, NDRG1 in 8q24.22 with translocation, RECQL4 in 8q24.3 with nonsense/frameshift/splice, TCEA1 in 8q11.23 with translocation, and TFEB in 6p21.1 with translocation (http://www.sanger.ac.uk/science/data/cancer-gene-census)(63) (Figure 14).

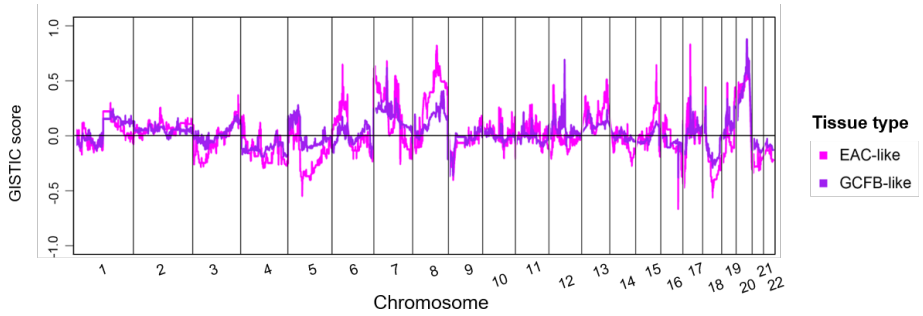


Figure 14. Copy number variation between EAC-like and GCFB-like in TCGA cohort.

In SNU cohort, after comparing putative genes with focal amplification by the GISTIC algorithm between EAC-like and GCFB-like, we identified 37 genes with significantly different copy number (*P*<0.05) (Table 5).

Table 5. Genes with significantly different copy number between EAC and GCFB in SNU cohort (P<0.05).

	Average Log_2CopyNumber in EAC-like	Average Log_2CopyNumber in GCFB-like
BOP1	0.332	0.026
C19orf12	0.279	0.039
DUSP8	0.267	0.002
EGFR	0.460	0.119
ERBB2	1.185	0.227
FOXP4	0.303	0.052
GRB7	0.826	0.219
GSTA1	0.252	0.001
GSTA2	0.320	-0.021
GSTA3	0.273	0.030
GSTA5	0.275	0.012
HIST1H1B	0.228	-0.051
HIST1H2AI	0.245	0.009
HIST1H2AK	0.234	-0.016
HIST1H2AL	0.282	-0.027
HIST1H2AM	0.281	-0.043
HIST1H2BM	0.215	-0.007
HIST1H2BN	0.230	-0.007
HIST1H2BO	0.239	-0.023

HIST1H3H	0.266	0.019
HIST1H3J	0.303	0.005
HIST1H4J	0.254	0.017
LILRA3	0.276	-0.123
LOC100287704	0.399	-0.011
LY86	0.236	-0.045
MDFI	0.381	0.047
MIEN1	1.178	0.205
OR2B2	0.209	-0.019
PI4KAP1	0.284	-0.010
PLEKHF1	0.343	0.048
POP4	0.289	0.053
SSR1	0.216	-0.002
TFEB	0.431	0.075
TMEM191B	0.381	-0.013
TRAM2	0.293	0.041
UGT2B17	0.263	-0.049
VSTM2B	0.270	0.069
ZNF439	-0.242	0.023

Out of those 37 genes, filtration using human Cancer Gene Census revealed that 2 genes, ERBB2 in 17q12 with amplification and TFEB in 6p21.1 with translocation, were selected as cancer related genes (Figure 15).

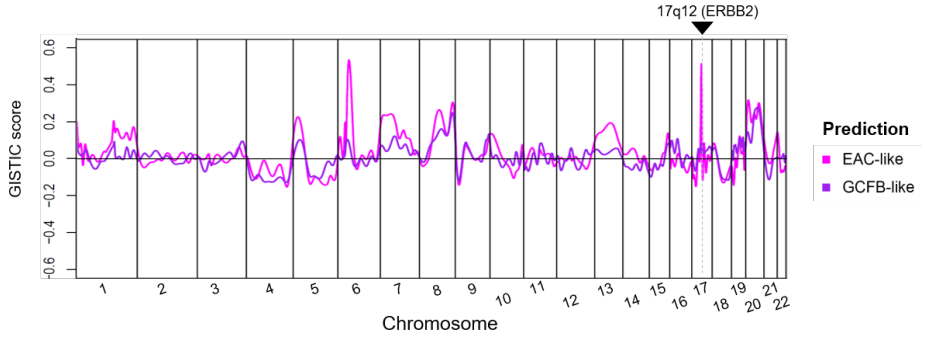


Figure 15. Copy number variation between EAC-like and GCFB-like in SNU cohort.

ERBB1 (EGFR) in 7p11.2 was focal amplified gene in both EAC-like and GCFB-like group simultaneously, but copy number of EGFR was not significantly different between 2 groups in SNU cohort. Because annotated mutation pattern of COX6C, HNRNPA2B1, NDRG1, RECQL4, TCEA1, and TFEB from both cohorts were inconsistent to copy number amplification, ERBB2 and ERRB1 as its possible heterodimer were validated using RPPA of TCGA cohort and tissue microarray of SNU cohort.

Protein expression of Reverse phase protein array and tissue microarray

Through supervised analysis of RPPA data comprised of 44 EAC and 88GCFB in TCGA cohort, we observed clearly separated clusters of expression with 81 proteins between EAC and GCFB proteins (Fig 16).

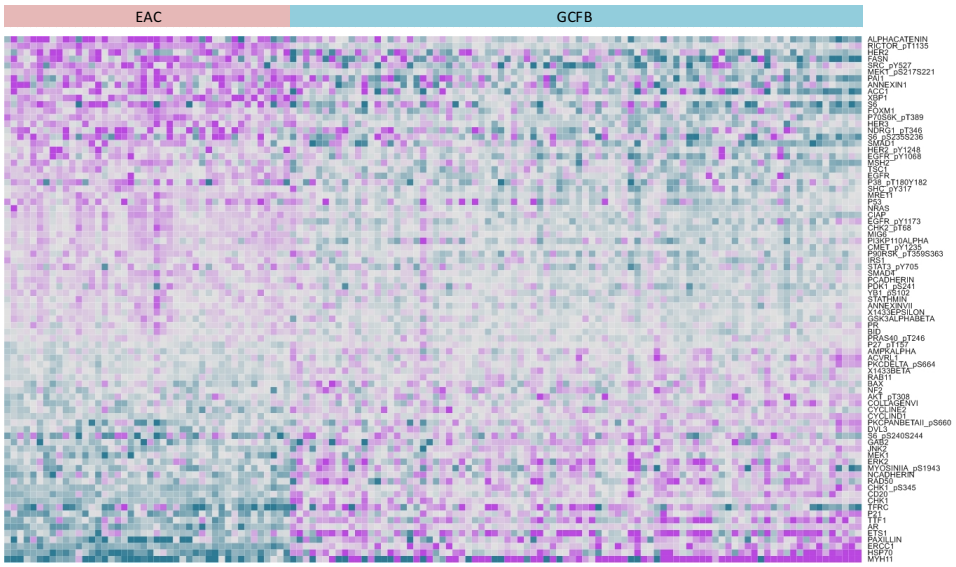


Figure 16. Heatmap using reverse phase protein assay of TCGA cohort.

Out of these 81 proteins, PIK3CA and AKT1 from pathway analysis of 400 signature genes, ERBB2 and EGFR from copy number analysis showed significantly different protein expression of RPPA between EAC and GCFB. For external validation, we analyzed different expression of these 4 proteins using 3 sets of TMA of SNU cohort with commercially available antibodies (Table 6).

Table 6. Information of antibodies for tissue microarray

Antibody	Clonality	Dilution	Detection kit	source	Cat. no
EGFR	Mouse	Ready	OptiView	Roche	790-

	monoclonal	to use	polymer (Ventana)		2988
ERBB2	Rabbit monoclonal	Ready to use	OptiView polymer (Ventana)	Ventana medical systems	790-2991
PI3Kinase p110alpha	Rabbit monoclonal	1:100	OptiView polymer (Ventana)	Cell signaling	#4249
AKT1	Rabbit monoclonal	1:50	OptiView polymer (Ventana)	Abcam	ab32505

The staining patterns of EGFR, ERBB2, PI3Kinasep110alpha, AKT1 in TMA are shown in Figure 17.

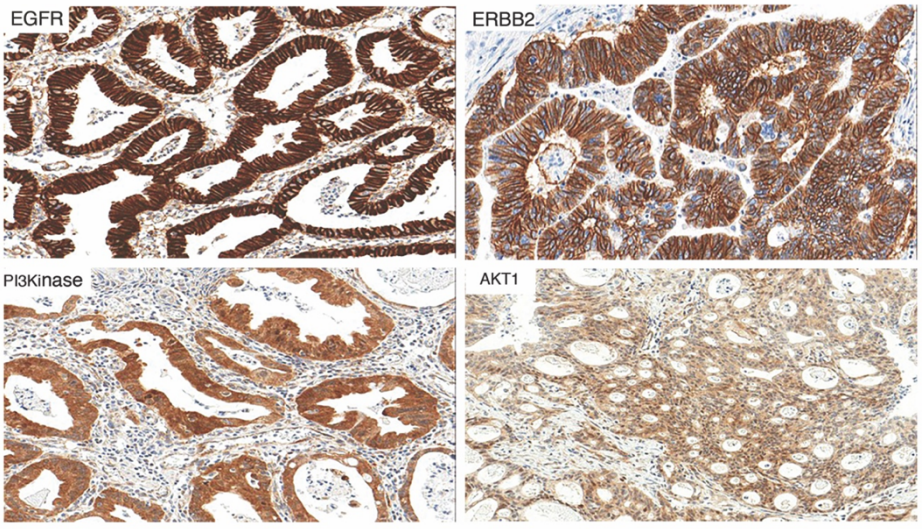


Figure 17. Protein expression using immunohistochemical staining of

tissue microarray (200x). EGFR, ERBB2, PI3Kinase showed staining of 3+ positivity and AKT1 showed up to 2+ positivity.

We calculated complex H score of EGFR, PI3Kinasep110alpha, AKT1 using expression results for each 3-different set of TMAs. Average H score of EGFR was significantly increased in EAC-like than in GCFB-like (160.7 ± 108.8 in EAC-like vs. 105.6 ± 81.6 in GCFB-like, P=0.014, Fig 18).

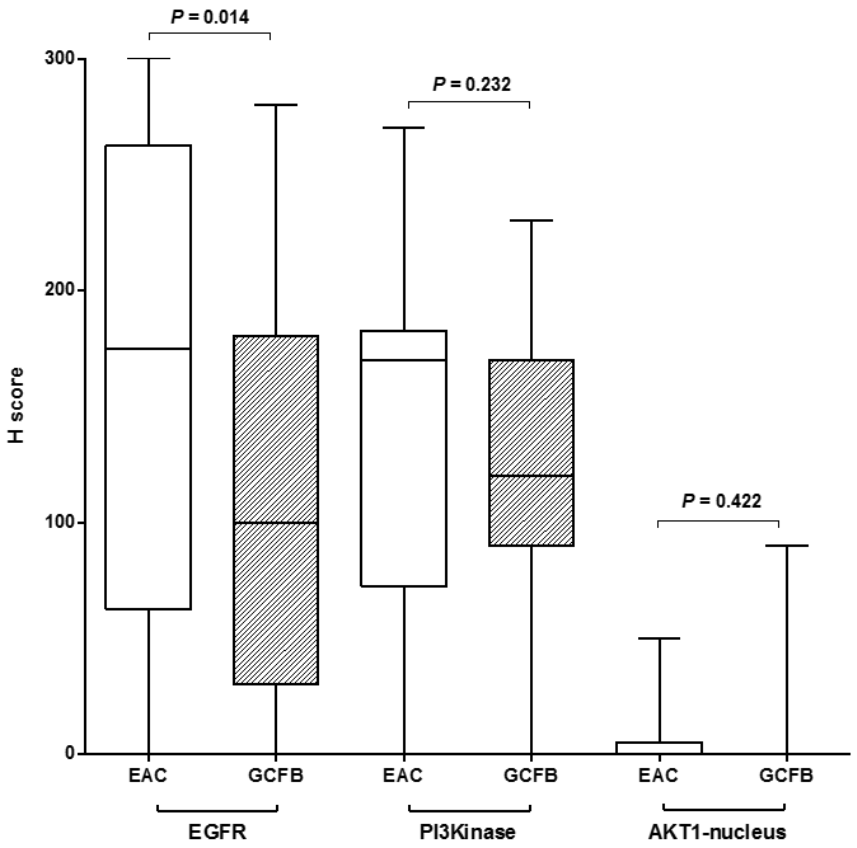


Figure 18. Complex H score of tissue microarray between EAC-like (n=10 x 3 sets) and GCFB-like (n=36 x sets) of SNU cohort.

However, there was no significant expression difference of PI3Kinase and AKT1. Staining results of IHC for ERBB2 revealed that ERBB2-positivity showed higher score tendency in EAC-like than GCFB-like (Table 7).

Table 7. Immunohistochemistry (IHC) and silver in situ hybridization (SISH) of ERBB2).

		EAC-like (n=10)	GCFB-like (n=36)	*P* value
IHC	0	3 (30.0%)	17 (47.2%)	
	1+	2 (20.0%)	14 (38.9%)	0.081
	2+	1 (10.0%)	2 (5.6%)	
	3+	4 (40.0%)	3 (8.3%)	
IHC and SISH	IHC<2+, or IHC 2+ and black/red ratio<2.0	5 (50.0%)	32 (88.9%)	0.015
	IHC 3+, or IHC 2+ and black/red ratio≧2.0	5 (50.0%)	4 (11.1%)	

Considering IHC and SISH together, EAC-like group shows significantly higher positivity (IHC 3+, or IHC 2+ and black/red ratio of SISH≧2.0) of ERBB2 compared to GCFB-like group (50.0% of EAC-

like vs. 11.1% of GCFB-like, *P*=0.015). All significant variables from univariate analysis in Table 3 were analyzed by multivariate analysis to identify risk factors for expression of EGFR and ERBB2. For overexpression of EGFR, prediction type (EAC-like or GCFB-like) was the only independent risk factor with 0.78 of adjusted R^2(*P*=0.034)(Table 8).

Table 8. Multivariate analysis for overexpression of EGFR.

Variable	Unstandardized coefficients B±standard error	Standardized coefficients β	*t*	*P* value	95% Confidence Interval for B
WHO classification	1.822±5.722	0.053	0.318	0.752	-9.733-13.378
Lauren classification	26.389±16.886	0.244	1.563	0.125	-7.665-60.443
Peri-neural invasion	-38.504±27.032	-0.220	-1.424	0.162	-93.057-16.049
Prediction type	62.500±28.509	0.314	2.192	0.034	5.044-19.956

For ERBB2 positivity, prediction type and WHO classification were independent risk factors (*P*=0.049 for prediction type and *P*=0.029 for

differentiated type)(Table 9).

Table 9. Multivariate analysis for ERRB2 positivity.

Variable		*P* value	Odds ratio	95% Confidence Interval for odds ratio
WHO classification (vs. undetermined)	differentiated	0.029	0.223	0.058-0.856
	undifferentiated	0.002	0.036	0.004-0.309
Lauren classification (vs. mixed)	intestinal	0.387	4.156	0.165-105.009
	diffuse	0.734	0.581	0.025-13.322
Perineural invasion (vs. invasion)	Non-invasion	0.576	0.532	0.058-4.870
Prediction type (vs.GCFB-like)	EAC-like	0.049	6.179	1.1011-37.752

External validation using CCLE database

We identified esophageal (n=3) and gastric (n=38) adenocarcinoma

cell lines with expression microarray data, SNP array data, and half maximal inhibitory concentration (IC50) for lapatinib, the dual EGFR and HER2 tyrosine kinase inhibitor, from CCEL database. Available data for each sample is presented in Table 10.

Table 10. Information of cell lines for esophageal and gastric adenocarcinoma from CCLE database.

Cell line	Organ	BCCP Score	prediction	Copy number of ERBB2*	Copy number of EGFR*	IC50†
OE33	Esophageal	0.546	EAC-like	amplification	0	3.538
OE19	Esophageal	0.402	GCFB-like	amplification	0	N/A
JHESOAD1	Esophageal	0.484	EAC-like	N/A	N/A	N/A
FU97	Gastric	0.109	GCFB-like	deletion	0	8.000
NUGC3	Gastric	0.37	GCFB-like	0	0	2.411
IM95	Gastric	0.318	GCFB-like	0	0	8.000
AGS	Gastric	0.19	GCFB-like	0	0	N/A
KATOIII	Gastric	0.536	EAC-like	0	0	N/A
SNU16	Gastric	0.351	GCFB-like	0	0	6.698
NCIN87	Gastric	0.753	EAC-like	amplification	0	0.066
OCUM1	Gastric	0.347	GCFB-like	0	0	8.000
SNU5	Gastric	0.291	GCFB-like	0	0	N/A
GCIY	Gastric	0.169	GCFB-like	0	0	7.255

SH10TC	Gastric	0.152	GCFB-like	0	0	8.000
MKN1	Gastric	0.341	GCFB-like	0	0	N/A
MKN74	Gastric	0.36	GCFB-like	0	amplification	4.690
KE39	Gastric	0.211	GCFB-like	amplification	0	4.056
HGC27	Gastric	0.062	GCFB-like	0	0	8.000
HUG1N	Gastric	0.315	GCFB-like	0	0	N/A
NUGC4	Gastric	0.313	GCFB-like	amplification	amplification	0.172
RERFGC1B	Gastric	0.365	GCFB-like	0	0	8.000
HS746T	Gastric	0.143	GCFB-like	0	0	8.000
NUGC2	Gastric	0.531	EAC-like	0	0	N/A
SNU1	Gastric	0.176	GCFB-like	0	0	8.000
MKN45	Gastric	0.341	GCFB-like	0	amplification	8.000
X2313287	Gastric	0.509	EAC-like	N/A	N/A	N/A
MKN7	Gastric	0.272	GCFB-like	amplification	0	8.000
SNU216	Gastric	0.37	GCFB-like	amplification	0	N/A

AZ521	Gastric	0.097	GCFB-like	0	0	1.660
LMSU	Gastric	0.129	GCFB-like	0	0	N/A
ECC10	Gastric	0.153	GCFB-like	0	0	N/A
TGBC11TKB	Gastric	0.326	GCFB-like	0	0	N/A
SNU520	Gastric	0.297	GCFB-like	0	0	N/A
GSS	Gastric	0.223	GCFB-like	0	amplification	N/A
SNU620	Gastric	0.322	GCFB-like	0	0	N/A
ECC12	Gastric	0.074	GCFB-like	0	0	N/A
GSU	Gastric	0.388	GCFB-like	0	0	N/A
SNU601	Gastric	0.507	EAC-like	0	0	N/A
SNU668	Gastric	0.144	GCFB-like	0	0	N/A
NCCSTCK140	Gastric	0.816	EAC-like	0	0	N/A
SNU719	Gastric	0.305	GCFB-like	0	amplification	N/A

*0 designates not-altered copy number and N/A not available.

†IC50 designates half maximal inhibitory concentration for lapatinib.

Using those cell lines, external validation using RNA microarray data of CCLE database showed significant difference of BCCP score between esophageal and gastric adenocarcinoma cell lines using Wilcoxon Rank Sum test (*P*=0.031)(Figure 19).

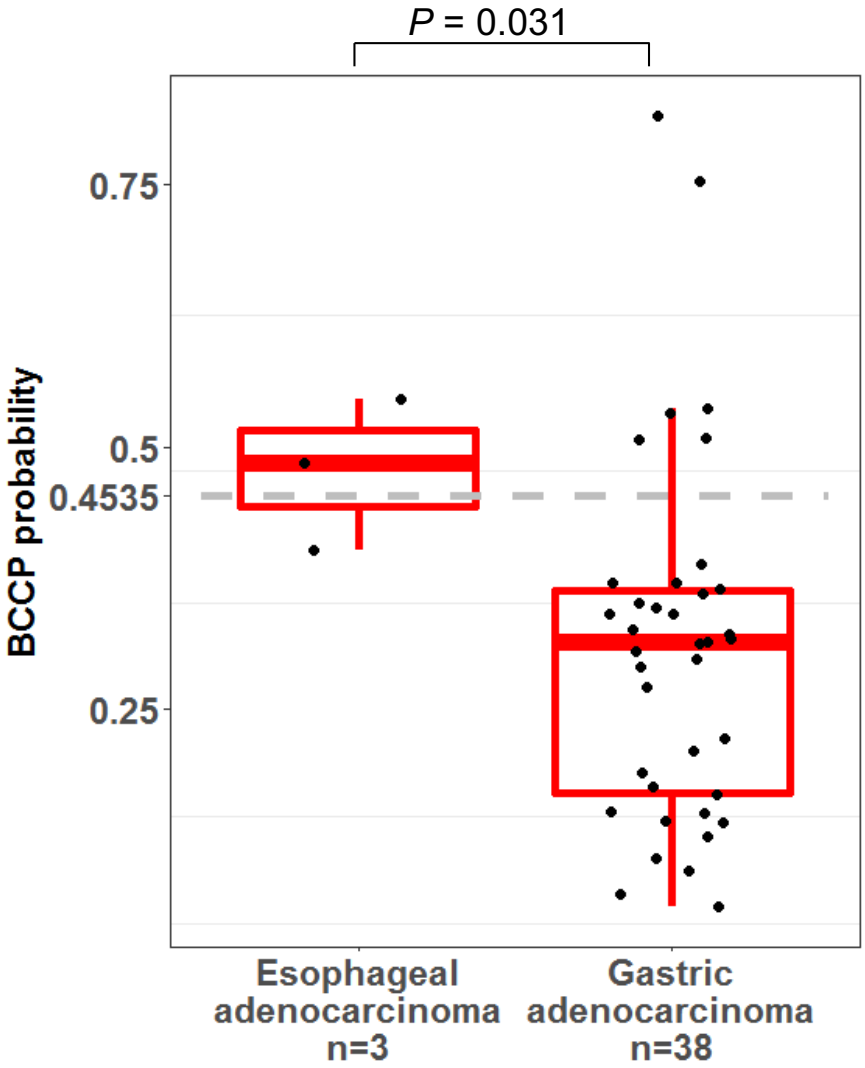


Figure 19. External validation of prediction model using CCLE database

Hierarchical clustering of CCLE database revealed that there was no

significant difference of tissue origin (Esophageal or gastric), ERBB2 amplification, or EGFR amplification between EAC-like and GCFB-like types using BCCP score (Figure 20).

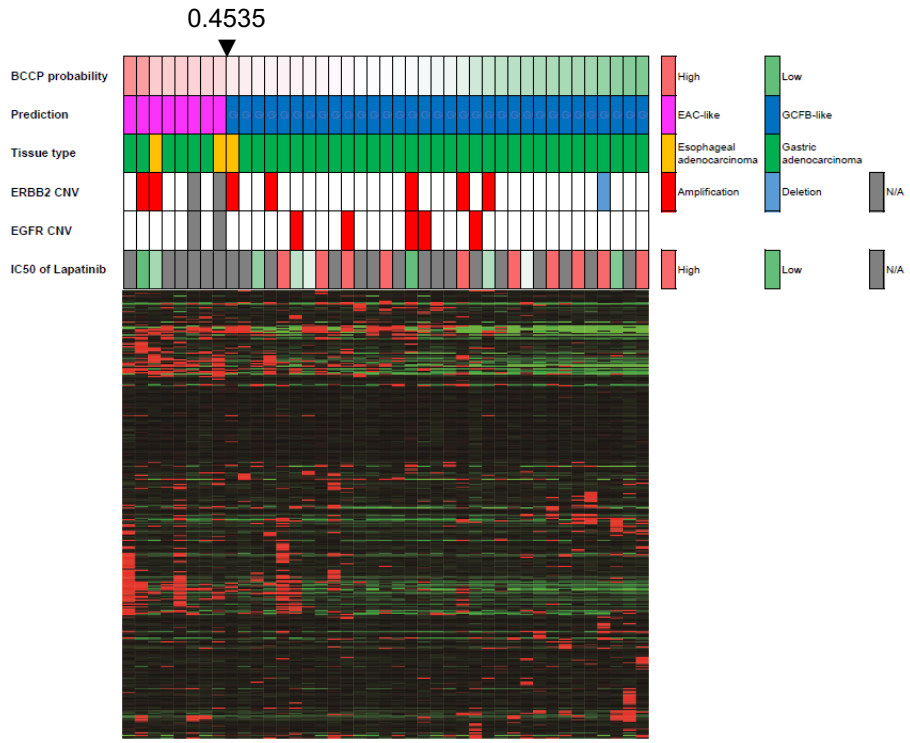


Figure 20. Hierarchical clustering of CCLE database between EAC-like and GCFB-like group

Target drug response of lapatinib, a dual EGFR and HER2 tyrosine kinase inhibitor, was evaluated using IC50 data of CCLE database between EAC-like (n=2) and GCFB-like groups (n=17)(Figure 21).

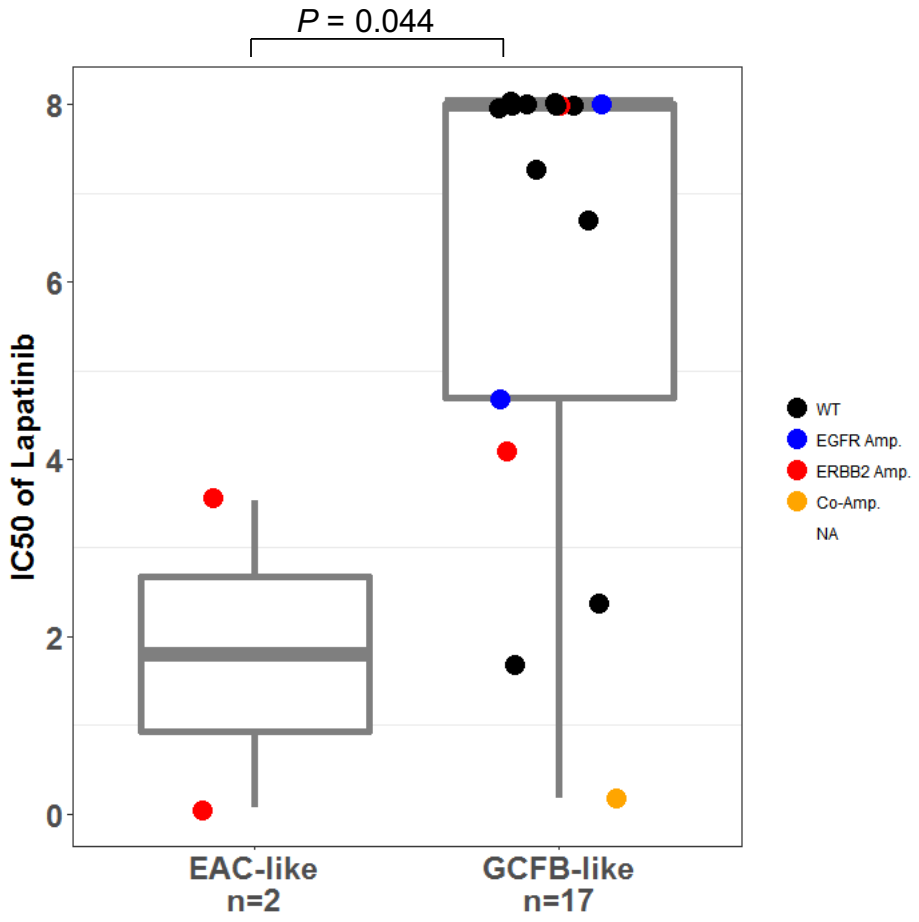


Figure 21. Drug response of lapatinib using half maximal inhibitory concentration (IC50) data of CCLE database between EAC-like and GCFB-like group

Analysis of IC50 demonstrated significantly lower IC50 for EAC-like than GCFB-like group using Wilcoxon Rank Sum test (*P*=0.044).

DISCUSSION

In this study, we successfully demonstrated molecular characteristics of AGEJ using next generation sequencing compared to pure esophageal or gastric adenocarcinoma, which presented a spectral transition of RNA expression between EAC-like and GCFB-like groups without any entirely distinguishable cluster. In addition, the same major proportion of AGEJ both in the East and the West, 68.8% of GEJ/Cardia in the West and of AGEJ II in the East, was classified as GCFB-like group. Interestingly, this geographic proportion of AGEJ (about 1/3 of EAC-like and 2/3 of GCFB-like) is similar to the proportion of the distance to oral (1cm) and aboral direction (2cm) between tumor epicenter and the gastroesophageal junction in conventional Siewert type II cancer(1, 64). This finding presumably represents that molecular classification from our study using the state-of-the-art analysis technique is consistent with that traditional geographic classification. For classification of AGEJ, especially Siewert type III, involvement of gastroesophageal junction by tumor has been an important criteria in traditional Siewert classification as well as AJCC TNM classification (1, 43). However, our previous study proposed that involvement of gastroesophageal junction be considered as a result of tumor progression and not related to an independent factor for classification of AGEJ in terms of postoperative prognosis(5). The current study also demonstrated that all AGEJ III involving gastroesophageal junction and most of AGEJ III without involving gastroesophageal junction were classified as GCFB-like group. Taken

together with our previous and current study, we could suggest that involvement of gastroesophageal junction is not a determinable factor to classify AGEJ III in terms of prognosis as well as molecular biology.

Tumor biology and geographic disparity of AGEJ has been well-known long-standing controversy between Eastern and Western institution. Traditionally, Siewert type I AGEJ is likely to have intestinal metaplasia or Barrett's esophagus, and gastroesophageal reflux or Barrett's mucosa has been known to be strong risk factors(65-67). Consequently, Siewert type I was usually considered and managed as a part of distal esophageal adenocarcinoma(3, 9). Siewert type III AGEJ is likely to show diffuse growth pattern with undifferentiated carcinoma and *H. pylori* infection could be significantly related to carcinogenesis, but possible inverse relationship to esophageal adenocarcinoma or Siewert type I cancer(65, 68, 69). As a result, Siewert type III was usually considered as a part of upper third gastric adenocarcinoma(4, 5, 65). However, the biologic relationship of both gastroesophageal reflux or *H. pylori* infection to Siewert type II, called as true GEJ cancer, was controversial (66, 68). Even there were a few studies proposing tumor biology of AGEJ as unique disease entity in terms of molecular analysis(35, 40, 70). Against this long-standing question, our study can propose that AGEJ is a certain biologic combination (approximately 1:2 proportion) of esophageal and gastric adenocarcinoma irrespective of the East or the West, not entirely similar to such one type of adenocarcinoma nor a completely distinctive entity.

Pathologically, previous studies suggested that there might be dichotomized carcinogenesis pathways of AGEJ consisted of intestinal metaplasia related pathway or non-intestinal pathway, but genetic relationship has not been proved (17, 71). In this study, we demonstrated that there was significant relationship of EAC-like group to intestinal type and GCFB-like group to diffuse type of previous studies. We expect that this consistent finding to previous pathologic reports will be promising supportive data for molecular analysis of intestinal metaplasia.

In this study, EAC group shows significantly increased copy number and protein overexpression of ERBB2. Anti-ERBB2 (HER2) monoclonal antibody, Trastuzumab, plus chemotherapy has been known to improve median overall survival significantly in patients with ERBB2-positive gastric/AGEJ cancer compared with chemotherapy alone(60). The positivity rate of ERBB2 was known as 22.1 % in gastric or gastroesophageal junction adenocarcinoma (61). Especially this positive rate was significantly higher in intestinal type (31.8 %) and gastroesophageal junction cancer (32.2 %) compared to diffuse type or other gastric cancer. Our data about EAC-like group was also significantly related to intestinal type and showed 50.0% of ERBB2 positivity which is much higher than previous report. On the other hands, GCFB-like group showed only 11.1% of ERBB2 positivity which is much lower than known positive rate of ERBB2 in usual gastric cancer or AGEJ. Considering this high positive rate of EAC-like group, we may

suggest that EAC-like adenocarcinoma by our molecular classification could be better indication for Trastuzumab treatment than usual gastric cancer or AGEJ. Interestingly, no ligand has been identified for ERRBB2 receptor which should dimerize (homo or hetero) with ligand-bound other members of ErbB receptor family for signal activation(72).

Epidermal growth factor receptor, or human epidermal growth factor receptor (HER1), is a member of the ErbB family of receptors that also includes HER2, HER3, and HER4 and a major partner for ERBB2 activation(73). EGFR ligand binding triggers the activation of downstream signaling tyrosine kinase pathways which control cell proliferation, survival, migration and also have a pivotal role during epithelial cell development in several organs(74-76). Regarding epithelial development, previous studies reported that elevated levels of EGFR have been identified in non-dysplastic intestinal metaplasia and may be involved in early event of the Barrett esophagus metaplasia, dysplasia, esophageal adenocarcinoma sequence (77-79). There previous studies are consistent with our results that EAC-like group in this study is significantly related to intestinal type and overexpression of EGFR. In the era of target therapy for cancer, recent several phase III randomized clinical trials reported that addition of most anti-EGFR antibodies including lapatinib, cetuximab, efitinib, or gefitnib to conventional chemotherapy failed to provide significant additional benefit for esophageal, gastric or AGEJ including Siewert type I and II adenocarcinoma (80-83). However, subgroup analysis of another

randomized clinical trial revealed that gefitinib could have advantage for selected esophageal adenocarcinoma or Siewert type I and II adenocarcinoma with EGFR amplification(84). According to the results of our study, about two-third or more of gastric adenocarcinoma or AGEJ II/III which were classified as GCFB-like group had significantly low protein expression of EGFR, and might become one possible explanation to show poor response to anti-EGFR antibodies in most previous clinical trials. On the contrary, we can expect that EAC-like group with significant amplification of ERBB2 and overexpression of EGFR would be a promising target for this new molecular treatment as a precision medicine. Moreover, because genes of most AGEJ and gastric cancer investigated in this study were found to be wild type, our molecular classification model is expected to be more promising tool not only for drug target of EAC-like adenocarcinoma but also designing new ERBB2 and EGFR-related clinical trial including EAC, AGEJ, and UT (85). Our study indirectly showed possibility of significantly different efficacy of lapatinib, a dual EGFR and HER2 tyrosine kinase inhibitor, according to genomic classification. Recently, novel pan-HER inhibitor, RB200, a bispecific (EGFR/HER3) ligand binding trap, was developed for a pan-HER therapy in human cancer(86). This pan-HER inhibitor inhibits phosphorylation of receptors in the HER family which results in several downstream signaling pathways, and also blocks EGFR/HER2, HER2/HER3, and HER3/HER4 heterodimer formation (87). In addition to ongoing phase III clinical trial for gastroesophageal cancer for

lapatinib, our data and future in-vivo validation based on genomic classification will be a promising evidence for novel target treatment for a subgroup of AGEJ (88).

We found similar expression of PI3Kinase and AKT between EAC-like and GCFB-like groups. This expression pattern of PI3Kinase and AKT was not consistent with pathway analysis using transcriptome expression which suggested PI3K-AKT pathway could be related to GCFB-like group. In EAC-like group, ligand binding of ERBB family has been known to trigger the activation of downstream signaling tyrosine kinase pathways including PI3K-AKT pathway also(73, 76). Therefore, we postulated that PI3K-AKT pathway could be controlled by both downstream activation of ERBB family in EAC-like group or overexpression of RNA clusters in GCFB-like group, which may eventually result in inconsistent protein expression pattern.

In conclusion, molecular profiling of AGEJ reveals that AGEJ consists of a combination of EAC-like and GCFB-like types characterized by 400 signature gene expression. Our newly developed predictive classification model demonstrated that GEJ/cardia in TCGA cohort and AGEJ II in SNU cohort were a combination of 31.2% of EAC-like group and 68.8% of GCFB-like group, not entirely similar to such one type of adenocarcinoma nor a completely distinctive entity. AGEJ III consisted of 93.7% of GCFB-like adenocarcinoma and there was no significant relationship between involvement of GEJ and molecular classification of AGEJ III. EAC-like group is significantly related to histological

differentiated and intestinal type, and GCFB-like group to undifferentiated and diffuse type, respectively. Compared to GCFB group, EAC group shows significantly increased copy number of ERBB2 and protein overexpression of ERBB2 and EGFR. We expect that our predictive model from comparable database of TCGA and SNU cohort could be useful classification system for esophageal, AGEJ and upper third gastric adenocarcinoma irrespective of epidemiologic difference in the future.

REFERENCES

1. Siewert JR, Holscher AH, Becker K, Gossner W. [Cardia cancer: attempt at a therapeutically relevant classification]. Chirurg. 1987;58(1):25-32.

2. Siewert J, Feith M, Werner M, Stein H. Adenocarcinoma of the esophagogastric junction: results of surgical therapy based on anatomical/topographic classification in 1,002 consecutive patients. Ann Surg. 2000;232(3):353-61.

3. Hulscher JB, van Sandick JW, de Boer AG, Wijnhoven BP, Tijssen JG, Fockens P, et al. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the esophagus. N Engl J Med. 2002;347(21):1662-9.

4. Sasako M, Sano T, Yamamoto S, Sairenji M, Arai K, Kinoshita T, et al. Left thoracoabdominal approach versus abdominal-transhiatal approach for gastric cancer of the cardia or subcardia: a randomised controlled trial. Lancet Oncol. 2006;7(8):644-51.

5. Suh YS, Han DS, Kong SH, Lee HJ, Kim YT, Kim WH, et al. Should adenocarcinoma of the esophagogastric junction be classified as esophageal cancer? A comparative analysis according to the seventh AJCC TNM classification. Ann Surg. 2012;255(5):908-15.

6. Japanese Gastric Cancer Association. Japanese gastric cancer treatment guidelines 2014 (ver. 4). Gastric Cancer. 2017 Jan;20(1):1-19.

7. Rice TW, Ishwaran H, Ferguson MK, Blackstone EH, Goldstraw P. Cancer of the Esophagus and Esophagogastric Junction: An Eighth Edition Staging Primer. J Thorac Oncol. 2017 Jan;12(1):36-42.

8. Hulscher JB, van Sandick JW, de Boer AG, Wijnhoven BP, Tijssen JG, Fockens P, et al. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the esophagus. N Engl J Med. 2002;347(21):1662-9.

9. Omloo J, Lagarde S, Hulscher J, Reitsma J, Fockens P, van Dekken H, et al. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the mid/distal esophagus: five-year survival of a randomized clinical trial. Annals of surgery. 2007;246(6):992-1000.

10. Kurokawa Y, Sasako M, Sano T, Yoshikawa T, Iwasaki Y, Nashimoto A, et al. Ten-year follow-up results of a randomized clinical trial comparing left thoracoabdominal and abdominal transhiatal approaches to total gastrectomy for adenocarcinoma of the oesophagogastric junction or gastric cardia. Br J Surg. 2015;102(4):341-8.

11. Yamashita H, Katai H, Morita S, Saka M, Taniguchi H, Fukagawa T. Optimal extent of lymph node dissection for Siewert type II esophagogastric junction carcinoma. Ann Surg. 2011;254(2):274-80.

12. Peng J, Wang WP, Yuan Y, Hu Y, Wang Y, Chen LQ. Optimal Extent of Lymph Node Dissection for Siewert Type II Esophagogastric

Junction Adenocarcinoma. Ann Thorac Surg. 2015;100(1):263-9.

13. Mine S, Sano T, Hiki N, Yamada K, Nunobe S, Yamaguchi T. Lymphadenectomy around the left renal vein in Siewert type II adenocarcinoma of the oesophagogastric junction. Br J Surg. 2013;100(2):261-6.

14. Hosoda K, Yamashita K, Katada N, Moriya H, Mieno H, Sakuramoto S, et al. Impact of Lower Mediastinal Lymphadenectomy for the Treatment of Esophagogastric Junction Carcinoma. Anticancer research. 2015;35(1):445-56.

15. Bang YJ, Kim YW, Yang HK, Chung HC, Park YK, Lee KH, et al. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. The Lancet. 2012;379(9813):315-21.

16. Sakuramoto S, Sasako M, Yamaguchi T, Kinoshita T, Fujii M, Nashimoto A et al. Adjuvant Chemotherapy for Gastric Cancer with S-1, an Oral Fluoropyrimidine. N Engl J Med. 2007 Nov 1;357(18):1810-20.

17. Nunobe S, Nakanishi Y, Taniguchi H, Sasako M, Sano T, Kato H, et al. Two distinct pathways of tumorigenesis of adenocarcinomas of the esophagogastric junction, related or unrelated to intestinal metaplasia. Pathology international. 2007;57(6):315-21.

18. Suh YS, Kong SH, Lee HJ, Yang HK. Reply to Letter:" Should Adenocarcinoma of the Esophagogastric Junction Be Classified as Gastric or Esophageal Cancer, or Else as a Distinct Clinical Entity?".

Ann Surg. 2016 Aug;264(2):e5-6.

19. von Rahden BH, Stein HJ, Feith M, Becker K, Siewert JR. Lymphatic vessel invasion as a prognostic factor in patients with primary resected adenocarcinomas of the esophagogastric junction. J Clin Oncol. 2005 Feb 1;23(4):874-9.

20. Cameron AJ, Lomboy CT, Pera M, Carpenter HA. Adenocarcinoma of the esophagogastric junction and Barrett's esophagus. Gastroenterology. 1995;109(5):1541-6.

21. Tsuji N, Ishiguro S, Tsukamoto Y, Mano M, Kasugai T, Miyashiro I, et al. Mucin phenotypic expression and background mucosa of esophagogastric junctional adenocarcinoma. Gastric cancer. 2004;7(2):97-103.

22. Jemal A, Center MM, DeSantis C, Ward EM.Global Patterns of Cancer Incidence and Mortality Rates and Trends. Cancer Epidemiol Biomarkers Prev. 2010 Aug;19(8):1893-907.

23. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. Journal of the National Cancer Institute. 2005;97(2):142-6.

24. Liu K, Yang K, Zhang W, Chen X, Chen X, Zhang B et al. Changes of Esophagogastric Junctional Adenocarcinoma and Gastroesophageal Reflux Disease Among Surgical Patients During 1988-2012: A Single-institution, High-volume Experience in China. Ann Surg. 2016 Jan;263(1):88-95.

25. Kusano C, Gotoda T, Khor CJ, Katai H, Kato H, Taniguchi H et

al. Changing trends in the proportion of adenocarcinoma of the esophagogastric junction in a large tertiary referral center in Japan. J Gastroenterol Hepatol. 2008 Nov;23(11):1662-5.

26. Okabayashi T, Gotoda T, Kondo H, Inui T, Ono H, Saito D, et al. Early carcinoma of the gastric cardia in Japan: is it different from that in the West? Cancer. 2000;89(12):2555-9.

27. Hasegawa S, Yoshikawa T, Cho H, Tsuburaya A, Kobayashi O. Is Adenocarcinoma of the Esophagogastric Junction Different between Japan and Western Countries? The Incidence and Clinicopathological Features at a Japanese High-Volume Cancer Center. World J Surg. 2009;33(1):95-103.

28. Kodera Y, Yamamura Y, Shimizu Y, Torii A, Hirai T, Yasui K, et al. Adenocarcinoma of the gastroesophageal junction in Japan: relevance of Siewert's classification applied to 177 cases resected at a single institution. J Am Coll Surg. 1999;189(6):594-601.

29. Driessen A, Nafteux P, Lerut T, Van Raemdonck D, De Leyn P, Filez L et al. Identical cytokeratin expression pattern CK7+/CK20− in esophageal and cardiac cancer: etiopathological and clinical implications. Mod Pathol. 2004 Jan;17(1):49-55.

30. Kim M, Lee H, Yang H, Kim W. Clinicopathologic and protein expression differences between cardia carcinoma and noncardia carcinoma of the stomach. Cancer. 2005;103(7):1439-46.

31. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. Nature.

2000 Aug 17;406(6797):747-52.

32. Shah MA, Khanin R, Tang L, Janjigian YY, Klimstra DS, Gerdes H, et al. Molecular classification of gastric cancer: a new paradigm. Clinical Cancer Research. 2011;17(9):2693-701.

33. Peters CJ, Rees JRE, Hardwick RH, Hardwick JS, Vowler SL, Ong CAJ, et al. A 4-Gene Signature Predicts Survival of Patients With Resected Adenocarcinoma of the Esophagus, Junction, and Gastric Cardia. Gastroenterology. 2010;139(6):1995-2004.e15.

34. Tan IB, Ivanova T, Lim KH, Ong CW, Deng N, Lee J, et al. Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. Gastroenterology. 2011;141(2):476-85. e11.

35. Tanière P, Martel-Planche G, Maurici D, Lombard-Bohas C, Scoazec JY, Montesano R, et al. Molecular and clinical differences between adenocarcinomas of the esophagus and of the gastric cardia. The American journal of pathology. 2001;158(1):33-40.

36. Lagarde SM, ten Kate FJW, Richel DJ, Offerhaus GJA, van Lanschot JJB. Molecular prognostic factors in adenocarcinoma of the esophagus and gastroesophageal junction. Ann Surg Oncol. 2007;14(2):977-91.

37. Peters CJ, Rees JRE, Hardwick RH, Hardwick JS, Vowler SL, Ong CAJ, et al. A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. Gastroenterology. 2010;139(6):1995-2004.e15.

38. Wang G, Hu N, Yang HH, Wang L, Su H, Wang C, et al. Comparison of global gene expression of gastric cardia and noncardia cancers from a high-risk population in china. PloS one. 2013;8(5):e63826.

39. Zali H, Rezaei-Tavirani M, Vafaee R, Rezaei-Tavirani M. Gastric cardia adenocarcinoma pathway analysis. Gastroenterology and Hepatology from bed to bench. 2013;6(0):2013-6(Suppl.1):S11-S18.

40. Li-Chang HH, Kasaian K, Ng Y, Lum A, Kong E, Lim H, et al. Retrospective review using targeted deep sequencing reveals mutational differences between gastroesophageal junction and gastric carcinomas. BMC Cancer. 2015 Feb 6;15:32.

41. Bass AJ, Reynolds SM, Laird PW, Curtis C, Shen H, Weisenberger DJ, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202-9.

42. Cancer Genome Atlas Research Network, Analysis Working Group: Asan University, BC Cancer Agency, Brigham and Women's Hospital, Broad Institute, Brown University et al. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017 Jan 12;541(7636):169-175.

43. Edge S, Byrd D, Carducci M, Compton C. AJCC cancer staging manual. New York, NY: Springer; 2009.

44. Bosman FT, Carneiro F, Hruban RH, Theise ND. WHO classification of tumours of the digestive system. World Health

Organization; 2010.

45. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

46. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9.

47. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv e-prints [Internet]. 2013 March 1, 2013; 1303. Available from: http://adsabs.harvard.edu/abs/2013arXiv1303.3997L.

48. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

49. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. Analysis of gene expression data using BRB-array tools. Cancer Inform. 2007 Feb 4;3:11-7.

50. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. Journal of computational biology : a journal of computational molecular cell biology. 2002;9(3):505-11.

51. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct

subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A. 2003 Aug 19;100(17):9991-6.

52. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014 Jan;42(Database issue):D199-205.

53. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.

54. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

55. Consortium GP. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

56. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013;493(7431):216-20.

57. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-8.

58. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012;22(8):1525-32.

59. Mazières J, Brugger W, Cappuzzo F, Middel P, Frosch A, Bara I et al. Evaluation of EGFR protein expression by immunohistochemistry using H-score and the magnification rule: re-analysis of the SATURN study. Lung Cancer. 2013 Nov;82(2):231-7.

60. Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet. 2010 Aug 28;376(9742):687-97.

61. Van Cutsem E, Bang YJ, Feng-Yi F, Xu JM, Lee KW, Jiao SC et al. HER2 screening data from ToGA: targeting HER2 in gastric and gastroesophageal junction cancer. Gastric Cancer. 2015 Jul;18(3):476-84.

62. Ross DS, Zehir A, Cheng DT, Benayed R, Nafa K, Hechtman JF et al. Next-Generation Assessment of Human Epidermal Growth Factor Receptor 2 (ERBB2) Amplification Status: Clinical Validation in the Context of a Hybrid Capture-Based, Comprehensive Solid Tumor Genomic Profiling Assay. J Mol Diagn. 2017 Mar;19(2):244-254.

63. Futreal, PA, Coin, L, Marshall M, Down T, Hubbard T, Wooster R et al. A CENSUS OF HUMAN CANCER GENES. Nature Reviews. Cancer, 4(3), 177–183.

64. Siewert JR, Stein HJ. Classification of adenocarcinoma of the esophagogastric junction. Br J Surg. 1998;85:1457-9.

65. Siewert JR, Feith M, Stein HJ. Biologic and clinical variations of adenocarcinoma at the esophago-gastric junction: relevance of a topographic-anatomic subclassification. J Surg Oncol. 2005;90(3):139-46; discussion 46.

66. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med. 1999;340(11):825-831.

67. Shaheen N, Ransohoff D. Gastroesophageal reflux, Barrett esophagus, and esophageal cancer: scientific review. JAMA. 2002 Apr 17;287(15):1972-81.

68. Wu A, Crabtree J, Bernstein L, Hawtin P, Cockburn M, Tseng C, et al. Role of Helicobacter pylori CagA+ strains and risk of adenocarcinoma of the stomach and esophagus. International Journal of Cancer. 2003;103(6):815-21.

69. Chow WH, Blaser MJ, Blot WJ, Gammon MD, Vaughan TL, Risch HA, et al. An inverse relation between cagA+ strains of Helicobacter pylori infection and risk of esophageal and gastric cardia adenocarcinoma. Cancer Res. 1998;58(4):588-90.

70. El-Rifai We, Moskaluk CA, Harper JC, Petroni GR, Bissonette EA, Jones DR, et al. Genetic differences between adenocarcinomas arising in Barrett's esophagus and gastric mucosa. Gastroenterology. 2001;121(3):592-8.

71. Demicco EG, Farris AB, Baba Y, Agbor-Etang B, Bergethon K, Mandal R, et al. The dichotomy in carcinogenesis of the distal

esophagus and esophagogastric junction: intestinal-type vs cardiac-type mucosa-associated adenocarcinoma. Modern pathology. 2011;24(9):1177-90.

72. Harari D, Yarden Y. Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. Oncogene. 2000 Dec 11;19(53):6102-14.

73. Yarden Y. The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities. Eur J Cancer. 2001 Sep;37 Suppl 4:S3-8.

74. Miettinen PJ, Berger JE, Meneses J, Phung Y, Pedersen RA, Werb Z et al. Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor. Nature. 1995 Jul 27;376(6538):337-41.

75. Sibilia M, Wagner EF. Strain-dependent epithelial defects in mice lacking the EGF receptor. Science. 1995 Jul 14;269(5221):234-8.

76. Kallergi G, Agelaki S, Kalykaki A, Stournaras C, Mavroudis D, Georgoulias V. Phosphorylated EGFR and PI3K/Akt signaling kinases are expressed in circulating tumor cells of breast cancer patients. Breast Cancer Res. 2008;10(5):R80.

77. Jankowski J, Hopwood D, Wormsley KG. Flow-cytometric analysis of growth-regulatory peptides and their receptors in Barrett's oesophagus and oesophageal adenocarcinoma. Scand J Gastroenterol. 1992;27(2):147-54.

78. Brito MJ, Filipe MI, Linehan J, Jankowski J. Association of

transforming growth factor alpha (TGFA) and its precursors with malignant change in Barrett's epithelium: biological and clinical variables. International journal of cancer. 1995;60(1):27-32.

79. Clemons NJ, Phillips WA, Lord RV. Signaling pathways in the molecular pathogenesis of adenocarcinomas of the esophagus and gastroesophageal junction. Cancer biology & therapy. 2013;14(9):782-95.

80. Hecht JR, Bang Y-J, Qin SK, Chung HC, Xu JM, Park JO, et al. Lapatinib in Combination With Capecitabine Plus Oxaliplatin in Human Epidermal Growth Factor Receptor 2–Positive Advanced or Metastatic Gastric, Esophageal, or Gastroesophageal Adenocarcinoma: TRIO-013/LOGiC—A Randomized Phase III Trial. J Clin Oncol. 2015;34(5):443-51.

81. Satoh T, Xu R-H, Chung HC, Sun G-P, Doi T, Xu J-M, et al. Lapatinib plus paclitaxel versus paclitaxel alone in the second-line treatment of HER2-amplified advanced gastric cancer in Asian populations: TyTAN—a randomized, phase III study. J Clin Oncol. 2014;32(19):2039-49.

82. Lordick F, Kang YK, Chung HC, Salman P, Oh SC, Bodoky G et al. Capecitabine and cisplatin with or without cetuximab for patients with previously untreated advanced gastric cancer (EXPAND): a randomised, open-label phase 3 trial. Lancet Oncol. 2013 May;14(6):490-9.

83. Dutton SJ, Ferry DR, Blazeby JM, Abbas H, Dahle-Smith A,

Mansoor W, et al. Gefitinib for oesophageal cancer progressing after chemotherapy (COG): a phase 3, multicentre, double-blind, placebo-controlled randomised trial. Lancet Oncol. 2014;15(8):894-904.

84. Petty RD, Dahle-Smith A, Miedzybrodzka Z, Dutton SJ, Murray GI, Stevenson D, et al. Epidermal growth factor receptor copy number gain (EGFR CNG) and response to gefitinib in esophageal cancer (EC): Results of a biomarker analysis of a phase III trial of gefitinib versus placebo (TRANS-COG). J ClinOncol. 2014; 32(15):4016.

85. Hedner C, Borg D, Nodin B, Karnevi E, Jirström K, Eberhard J. Expression and Prognostic Significance of Human Epidermal Growth Factor Receptors 1 and 3 in Gastric and Esophageal Adenocarcinoma. PLoS One. 2016 Feb 4;11(2):e0148101.

86. Huang Z, Brdlik C, Jin P, Shepard HM. A pan-HER approach for cancer therapy: background, current status and future development. Expert Opin Biol Ther. 2009 Jan;9(1):97-110.

87. Nam HJ, Ching KA, Kan J, Kim HP, Han SW, Im SA et al. Evaluation of the antitumor effects and mechanisms of PF00299804, a pan-HER inhibitor, alone or in combination with chemotherapy or targeted agents in gastric cancer. Mol Cancer Ther. 2012 Feb;11(2):439-51.

88. Okines A, Cunningham D, Chau I. Targeting the human EGFR family in esophagogastric cancer. Nat Rev Clin Oncol. 2011 Apr 5;8(8):492-503.

국문 초록

서론: 위식도경계부선암의 생물학적인 이해 및 위선암 혹은 식도선암과의 분자생물학적 특징의 비교에 대해 끊임 없는 논란이 지속되어 왔다. 본 연구의 목적은 the Cancer Genome Atlas (TCGA) 와 서울대학교 코호트에 대한 차세대염기서열분석 유전체 데이터를 이용하여, 위식도경계부선암의 분자생물학적 특징을 식도선암 및 위선암과 비교 연구하고자 한다.

방법: TCGA 코호트로부터 식도선암(7 례), 위식도경계부/분문부 선암(48 례), 위저부/체부 선암(102 례)의 차세대염기서열분석 유전체 데이터를 추출하였다. 서울대학교 코호트로부터 위식도경계부선암 II 형(16 례), 위식도경계부선암 III 형(16 례), 상부위암(14 례)에 대해 종양 및 같은 환자의 정상 위점막 조직을 짝을 이루어(총 92 례) 전체 엑솜 및 전체 전사체 서열분석을 시행하였다. TCGA 코호트의 식도선암과 위저부/체부 선암 간의 전사체 데이터로부터 생성된 Bayesian compound covariate predictor (BCCP) 및 Leave-one-out cross validation 을 이용하여 분자생물학적 분류 예측 모델을 구축하고, TCGA 코호트의 위식도경계부/분문부 선암과 서울대학교 코호트의 전체 암종을 테스트하였다.

결과: 전사체 발현 차이를 보이는 400 개 유전자를 기반으로 분자생물학적 분류 예측 모델을 구축하여 적용 결과, TCGA 코호트의 위식

도경계부/분문부 선암 및 서울대학교 코호트의 위식도경계부선암 II 형의 68.8%는 BCCP 점수 0.4535 를 기준으로 위저부/체부양 선암으로 분류되었다(민감도 90.2%, 특이도 89.7%). 서울대학교 코호트의 위식도경계부선암 III 형의 93.7%는 위저부/체부양 선암으로 분류되었으며, 위식도경계부선암 III 형의 위식도경계부의 침범여부와 분자생물학적 분류 간에 유의한 연관성은 없었다. 위저부/체부양 선암에 비해 식도양 선암은 분화암 및 장형암과의 연관성이 유의하게 높았으며, ERBB2 의 유전자복제수변이도 유의하게 증폭되어 있었었다. Reverse phase protein array 와 조직미세배열법 결과, 식도양 선암에서 ERBB2 와 EGFR 의 단백질 발현이 유의하게 증가되어 있었다. Cancer Cell Line Encyclopedia 데이터베이스를 이용하여, EGFR 과 ERBB2 의 이중티로신키나제억제제인 lapatinib 에 대한 약물 반응성 분석 결과 위저부/체부양 선암에 비해 식도양 선암에서 유의하게 낮은 IC50 수치를 확인하였다.

결론: 400 개 유전자를 기반으로 분자생물학적 분류 예측 모델을 구축하여 적용한 결과, TCGA 코호트의 위식도경계부/분문부 선암 및 서울대학교 코호트의 위식도경계부선암 II 형은 31.2%의 식도양 선암과 68.8%의 위저부/체부양 선암의 조합으로 이루어져 있었다. 식도양 선암은 분화암 및 장형암과의 연관성이 유의하게 높았으며, ERBB2 의 유전자복제수변이가 유의하게 증폭되어 있고, ERBB2 및 EGFR 의 단백질 발현이 유의하게 증가되어 있었다. 식도양선암은

ERBB2 및 EGFR 티로신키나제억제제에 대한 유망한 표적이 될 수 있을 것으로 기대된다.

주요어 : 위식도경계부선암, 위선암, 식도선암, 차세대염기서열분석, **유전체**

학 번 : 2012-31122